The demographics and evolution of the absorbed AGN population

Tom Dwelly

Mullard Space Science Laboratory
Department of Space and Climate Physics
University College London

A thesis submitted to the University of London
for the degree of Doctor of Philosophy

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

It has become increasingly apparent that active galactic nuclei (AGN) have played a key role in the galaxy formation process, leading to the galaxy population we see today. In order to understand better this inter-relationship, we must first measure the characteristics and evolution of the AGN population over cosmic timescales. Models of the AGN population which reproduce the spectrum and intensity of the extragalactic X-ray background require a large population of faint AGN, the majority obscured by large column densities of cold material. In the local Universe, where we find mostly low luminosity objects, the obscured AGN make up $\sim 80\%$ of the population. However, at higher redshifts, including the epoch when AGN and galaxies were forming most rapidly, the demographics of the obscured AGN population are still poorly understood. For this thesis, I have made a detailed examination of the AGN detected in several of the deepest *XMM-Newton* "blank-field" observations. I have carried out a detailed set of Monte-Carlo simulations in order to compare directly the X-ray properties of the observed AGN to the predictions of a number of AGN population models. For sources detected in the "13^{H}" deep field, I find that the best fitting model contains AGN with a broad range of obscuration levels, but with significantly absorbed AGN making up at least 75% of the population. Furthermore, by examining the sources in *XMM-Newton* observations of the "CDFS" field, for which nearly complete redshift determinations are available, I find that the AGN absorption distribution exhibits little redshift or luminosity dependence. I confirm these findings by extending my study to a much larger AGN sample, and investigate field-to-field AGN source density variations.

Contents

List of Figures 8

Acknowledgements 13

1 Introduction 14

1.1 Motivation 14

1.2 The unified scheme and the AGN menagerie 16

1.2.1 The central engine and X-ray emission from AGN 17

1.2.2 X-ray absorption and scattering processes 19

1.2.3 Optical classes of AGN, and their spectral properties 26

1.2.4 Broad absorption line quasars 31

1.3 Charting the AGN population 36

1.3.1 Extragalactic X-ray surveys 37

1.3.2 AGN surveys at other wavelengths 40

1.4 The demographics and evolution of the AGN population 42

1.4.1 The AGN luminosity function 42

1.4.2 The distribution of absorption in the AGN population 46

1.4.3 How can the AGN population be better constrained? 47

1.4.4 The influence of large scale structure on deep surveys 51

1.5 The layout of this thesis 52

2 Data reduction and Monte Carlo simulations 53

2.1 *XMM-Newton* EPIC data reduction 53
2.1.1 Processing raw *XMM-Newton* EPIC data 54
2.1.2 Minimising background effects 54
2.1.3 The EPIC energy bands used in this study 55
2.1.4 Generating images and exposure maps 57
2.1.5 Astrometric precision of the EPIC detectors 57
2.1.6 Additional considerations 58
2.1.7 Conversion from EPIC count rate to flux 59
2.1.8 Background fitting and source detection 60
2.2 Monte Carlo simulations of *XMM-Newton* EPIC observations 65
2.2.1 Modelling the *XMM-Newton* point spread function 65
2.2.2 Making simulated EPIC images 66
2.2.3 Modelling the AGN population 69

3 Constraints on the distribution of absorption using X-ray-selected AGN in the 13^{H} deep field 71

3.1 Introduction and motivation 71
3.2 *XMM-Newton* observations in the 13^{H} field 73
3.3 Source detection and sourcelist cleaning 75
3.3.1 Input source populations 76
3.3.2 Simulated *XMM-Newton* images 78
3.3.3 Matching of input sources and output detections 79
3.3.4 Determining the detection likelihood thresholds 80
3.3.5 Removing non-AGN sources from the sample 83
3.4 Modelling the AGN population 84
3.4.1 Models for the distribution of absorption in AGN 85
3.4.2 Modelling the distribution of X-ray spectral slopes 89
3.4.3 The X-ray luminosity function 89
3.4.4 X-ray colours from AGN spectral templates 94

3.4.5 Carrying out the Monte Carlo simulations 96
3.5 Capabilities of the 13^H survey 96
3.5.1 X-ray selection function 97
3.5.2 Sensitivity of X-ray colours to absorption 99
3.5.3 X-ray colour distributions produced by model AGN populations 103
3.6 Results 105
3.6.1 Colour distribution of the 13^H sample 105
3.6.2 Reproducing the source counts in the 13^H sample 106
3.6.3 Statistical comparison of observed and predicted colour distributions 108
3.6.4 Calculating 3D-KS test probabilities 110
3.7 Selecting absorbed sources from their X-ray colours 111
3.8 Discussion of results 114
3.8.1 Reproducing the X-ray colours in the 13^H sample 114
3.8.2 The X-ray colours of the "hard" sources 116
3.8.3 Implications for AGN torus models 117
3.8.4 Source count disparity 118
3.8.5 High redshift AGN in the 13^H field 120
3.9 Summary 120
3.10 Extending the analysis to other fields 121

4 The AGN population in the CDFS: the *XMM-Newton* view 123
4.1 Introduction 123
4.2 Observations and data reduction 125
4.2.1 The *XMM-Newton* dataset 125
4.2.2 *XMM-Newton* source detection procedure 126
4.3 Matching to *Chandra* and optical catalogues 129
4.3.1 Cross correlation with *Chandra* observations of the field 129
4.3.2 Source confusion in the XMM-CDFS sample 131

4.3.3 Optical counterparts and redshifts 133
4.3.4 A summary of the XMM-CDFS sample 137
4.4 Models of the AGN N_H distribution 137
4.5 Estimating the intrinsic properties of the sample using X-ray colours and Monte Carlo simulations 140
4.5.1 Constructing the library of simulated sources 142
4.5.2 Absorption estimation technique 146
4.5.3 Intrinsic luminosity estimation technique 147
4.5.4 Fidelity of the N_H/L_X estimation technique 147
4.5.5 X-ray completeness of the XMM-CDFS sample 154
4.5.6 Positional accuracy of the *XMM-Newton* detections in the XMM-CDFS 154
4.5.7 Direct comparison of the XMM-CDFS sample with model AGN populations 156
4.6 Results 157
4.6.1 Applying the N_H/L_X estimation technique to the XMM-CDFS sample 157
4.6.2 Source counts in the XMM-CDFS 158
4.6.3 The redshift and luminosity distributions in the XMM-CDFS sample 158
4.6.4 The N_H distribution in the XMM-CDFS sample 162
4.6.5 The absorbed fraction in the XMM-CDFS and its dependence on luminosity and redshift 162
4.6.6 Distribution of the XMM-CDFS sample in z,L_{2-10},N_H space . 165
4.7 Discussion of results 168
4.7.1 Ability of AGN population models to reproduce the absorption distribution in the XMM-CDFS sample 168
4.7.2 The XMM-CDFS sources without redshift determinations . . 170
4.7.3 XMM-CDFS sources with unusual/complex spectra 173

4.8 Regions of extended emission in the *XMM-Newton* observations of the CDFS 176
4.9 Comparison to the *Chandra* spectral analysis of Tozzi et al. (2006) . . 176
4.9.1 Why have other X-ray surveys arrived at different conclusions? 178
4.9.2 Large scale clustering in the CDFS 182

5 The *XMM-Newton* Deep One Degree Survey: A large sample of faint X-ray selected AGN 185
5.1 Introduction 185
5.2 Data reduction method for individual *XMM-Newton* fields 186
5.2.1 The 1^{H} field 187
5.2.2 The 13^{H} field revisited 192
5.2.3 The CDFN field 192
5.2.4 The Lockman Hole field 192
5.2.5 The LBQS 2212 field 195
5.2.6 The CDFS field revisited 197
5.2.7 Summary of the XMM-DODS sample 197
5.3 Modelling the XMM-DODS sample 198
5.4 X-ray colour distribution in the XMM-DODS sample 200
5.4.1 Comparison to the X-ray colour distributions predicted by AGN population models 203
5.5 Source count distributions in the XMM-DODS sample 209
5.5.1 Defining single-band XMM-DODS samples 211
5.5.2 The sky area covered by the XMM-DODS observations 211
5.5.3 Reproducing the source counts in the XMM-DODS sample . . 214
5.5.4 Field to field density variations as an indicator of large scale clustering 219
5.6 Discussion of the XMM-DODS results 224
5.6.1 X-ray colour distributions 224

5.6.2 Source count distributions 225
5.6.3 Field to field source density variations 226

6 Discussion and Conclusions **227**
6.1 Tracing the absorbed AGN population 227
6.2 The luminous absorbed AGN population 228
6.3 Obscured black hole growth over cosmic timescales 229
6.4 Summary 231
6.5 Future directions and outlook 231

A X-ray colour-colour diagnostic plots **234**

B Reduction of Keck-LRIS spectra **236**
B.1 Keck-LRIS observations 236
B.2 Reduction process 237

C An optically faint, X-ray bright source in the XMM-CDFS **240**

Bibliography **243**

List of Figures

1.1 Cartoon schematic of the inner regions of an AGN 18
1.2 AGN spectra: absorbed power law 21
1.3 AGN spectra: ionised absorber 22
1.4 AGN spectra: reflection component 24
1.5 Optical spectra of BLAGN 27
1.6 Optical spectra of NELGs 29
1.7 Optical spectra of a type-2 QSO 30
1.8 Optical spectra of BALQSOs 32
1.9 Spectrum of the XRB 48

2.1 Particle background rejection for an EPIC observation 56
2.2 Vignetted and unvignetted background components 64
2.3 Point spread function model for the EPIC detectors 67

3.1 13^{H} *XMM-Newton* image 74
3.2 Flow chart of the simulation process 77
3.3 Flux measurement reliability 81
3.4 Detection likelihood thresholds 82
3.5 Model N_H distributions 88
3.6 The LDDE1 XLF model of Miyaji et al. (2000) 91
3.7 13^{H} field X-ray selection function 98
3.8 Dependence of the X-ray selection function on N_H and Γ 99

3.9 Sensitivity of X-ray colours to absorption 102
3.10 Observed and predicted colour-colour distributions 104
3.11 X-ray colour-colour distributions found in the 13^{H} sample 105
3.12 Observed and predicted $N(> S_{0.5-2})$ curves 107
3.13 Selecting absorbed sources using hardness ratios 111

4.1 XMM-CDFS X-ray images 127
4.2 Multi-wavelength coverage of the XMM-CDFS 128
4.3 *XMM-Newton/Chandra* position differences 132
4.4 Sources with incorrect optical counterparts in (Zheng et al., 2004) . . 135
4.5 X-ray/optical position differences 138
4.6 X-ray flux versus optical magnitude 139
4.7 The Ueda et al. (2003) XLF model 144
4.8 Fidelity of the N_H estimation technique 149
4.9 Fidelity of the luminosity estimation technique 150
4.10 Fidelity of the N_H estimation technique 151
4.11 Fidelity of the N_H/L_X estimation process 153
4.12 X-ray Completeness the XMM-CDFS observations 155
4.13 Differential source counts in the XMM-CDFS 159
4.14 The redshift and luminosity distributions of the XMM-CDFS sample 160
4.15 N_H distribution in the XMM-CDFS sample 161
4.16 The absorbed fraction in the XMM-CDFS 164
4.17 Distribution of absorption *vs* intrinsic luminosity 166
4.18 Distribution of absorption *vs* redshift 167
4.19 Hardness ratios 171
4.20 Comparison with Tozzi et al. (2006) 177
4.21 Effect of not correcting luminosities for absorption 181
4.22 B magnitude *vs* redshift in the CDFS 183

5.1 *XMM-Newton* images of the XMM-DODS fields 190

5.2 PSF flux mismatches in the Lockman Hole 196
5.3 Distribution of flux ratios in the XMM-DODS sample 202
5.4 Flux ratio distributions predicted by the N_H models 204
5.5 Observed and predicted cumulative flux ratio distributions 206
5.6 Observed and predicted differential flux ratio distributions 210
5.7 Differential sky area of the six XMM-DODS fields 212
5.8 Differential sky area of the XMM-DODS observations 213
5.9 Differential source counts (Ueda et al. 2003 XLF model) 217
5.10 Differential source counts (Miyaji et al. 2000 LDDE1 XLF model) . . 218
5.11 Integral source counts in the six XMM-DODS fields 222
5.12 Differential source counts in the six XMM-DODS fields 223

A.1 HR and FR as a function of AGN parameters 235

C.1 Multi-wavelength images of a QSO-2 candidate 242

List of Tables

2.1 *XMM-Newton* EPIC energy conversion factors 61
2.2 Relative response of different EPIC detector/filter combinations . . . 61

3.1 Parameters for the LDDE1 XLF model of Miyaji et al. (2000) 92
3.2 Kolmogorov-Smirnov test results 109
3.3 Observed and predicted "hard" fraction 113

4.1 Parameters for the XLF model of Ueda et al. (2003) 143
4.2 Statistical comparisons of the XMM-CDFS sample and the predictions of the N_H models 163

5.1 Summary of *XMM-Newton* observations 188
5.2 Summary of the XMM-DODS fields 191
5.3 Statistical comparison of the XMM-DODS source count distributions with model predictions 216

Acknowledgements

Above all I am indebted to Mat Page for all the time and effort he has invested to guide me to this stage. I'd have never got this thing finished without his knowledge, enthusiasm, patience, humour and, maybe most importantly, his tea-maker. I couldn't have asked for a better PhD supervisor. Cheers Mat!

Many thanks to my secondary supervisor Graziella Branduardi-Raymont for going through my thesis with a fine-tooth comb, for making very useful suggestions, and for asking me unpleasant questions in panel meetings. Thanks also to all of the other various MSSLites who along the way have helped me out with scientific and technical advice, and of course friendship. Being a part of the ongoing $13^{H}/1^{H}$ deep field project has been very productive, so thanks to all of my collaborators. I'm especially grateful for being sent off on so many observing trips to exotic locations. Long may it continue! Thanks as well to my PhD examiners Omar Almaini and Steve Warren for an interesting (and fairly painless) viva, and for making excellent suggestions for future extensions to the work presented in this thesis.

Last but not least, muchos gracias to Patricia, Mum, Dad and Rosie for motivating me through the dark days of the thesis write up, and for providing assorted sheds in which to store my rubbish.

Thanks again to all who have helped.

Chapter 1

Introduction

1.1 Motivation

Active galactic nuclei (AGN) are the most consistently luminous objects in the Universe; the brightest outshine the starlight from their host galaxies by many orders of magnitude. However, AGN are far from a homogeneous population: they span many orders of magnitude in luminosity, exhibit a variety of spectral characteristics, and have been detected at all epochs since the Universe was $\leq 5\%$ of its current age. Historically, astronomers have grouped AGN into various classes, based on their prime means of detection. For example, powerful, distant quasars were first detected from their radio emission, whereas the presence of AGN in nearby Seyfert galaxies was revealed by their prominent optical emission lines. However, it is thought that all AGN are driven by a common power source; namely, accretion of matter onto a super-massive black hole residing at the centre of a host galaxy. The wide range of luminosities can be explained if the AGN have a range of black hole masses and accretion rates. However, other observed differences cannot be explained by simple scaling of the central engine. In response, "unification" schemes have been devised which attribute the wide variety of multi-wavelength properties seen in AGN to the geometry and relative orientation of gas and dust surrounding the central black hole

(e.g. Antonucci, 1993).

The high luminosities of AGN mean that they can be detected to immense distances. It has therefore been possible to chart their cosmological history, revealing that the peak in AGN activity occurred when the Universe was about a third of its current age ($1 < z < 2$). During this epoch, the co-moving AGN luminosity density was around 100 times greater than that seen in the local ($z < 0.1$) Universe (e.g. Page et al., 1997; Miyaji et al., 2000). The total energy output from AGN over cosmic history is thought to account for as much as 20% of that emitted as starlight (Hasinger, 2000). What is more, the growth of super-massive black holes is closely tied to the building of the stellar bulges in their host galaxies (e.g. Magorrian et al., 1998). AGN activity must somehow be intertwined with the processes which have built the stars we see in massive galaxies today. However the nature of this relationship is poorly understood; is star formation driven by AGN activity? Or do other processes which initiate star formation also fuel black hole growth? Can very luminous AGN curtail star formation by heating the gas in the host galaxy which otherwise would have collapsed to form stars? Models which attempt to answer these questions cannot be fully tested until we have a much better understanding of the cosmic history of AGN activity.

A seemingly ubiquitous feature of AGN is prodigious X-ray emission. A few percent of the bolometric luminosity of quasars is emitted in the X-ray band (e.g. Elvis et al., 1994; Barger et al., 2005). This means that X-ray surveys are an efficient way to detect large numbers of AGN. Progressively deeper "blank field" X-ray surveys have detected greater and greater sky densities of point-like sources, the vast majority of which are AGN. The integrated emission from the AGN detected in the deepest X-ray surveys (e.g. Giacconi et al., 2002; Alexander et al., 2003), can account for more than 80% of the intensity of the 0.5–2 keV extragalactic X-ray background (XRB) first detected by Giacconi et al. (1962).

However, there is an apparent paradox; the spectrum of the XRB in the 1–30 keV range is much harder (photon flux $\propto E^{-1.4}$) than the X-ray spectra of AGN

typically detected in soft X-ray surveys (photon flux $\propto E^{-2}$). Heavily obscured AGN are therefore thought to emit the bulk of the hard XRB. Hard X-rays are able to penetrate through substantial column densities of obscuring material that are effectively opaque to optical, ultraviolet (UV) and soft X-ray emission. In the local Universe, around 80% of the nuclei of Seyfert galaxies show substantial obscuration in their X-ray spectra. However, at higher redshifts, the distribution of absorption in the AGN population is poorly understood; the absorbed AGN population remains largely unconstrained. Therefore, many questions are still open. Does the pattern of absorption seen in the low redshift, low luminosity Seyferts extend to the high redshift, high luminosity quasar regime? Are absorbed AGN more/less common in the local Universe than at earlier times? Are the obscured AGN intrinsically similar systems to unobscured AGN (as suggested by the unification schemes), or are they actually very different objects? Did the peak in obscured AGN activity occur at the same epoch ($z \sim 2$) as the peak in unobscured quasar activity?

The properties of individual AGN in the local Universe can be studied in detail because of their apparent brightness. However, if we are to understand the high-z AGN population, particularly the obscured objects, then large numbers of faint AGN must be studied. For this we require deep X-ray surveys.

In this thesis I present the results of my efforts toward a better understanding of the demographics and evolution of the obscured AGN population. I will firstly provide an introduction to some AGN processes, the methods that have been used to detect AGN, and how their population is thought to have evolved over cosmic timescales.

1.2 The unified scheme and the AGN menagerie

A widely accepted model has emerged to explain the observations of AGN at all wavelengths. It is thought that the radiative output in all types of AGN is driven by accretion of matter onto a super-massive black hole. No other reasonable mechanism

is able to explain such a concentrated production of radiation. This accretion process drives, either directly or indirectly, all of the multi-wavelength spectral features we see from AGN.

However, the picture is complex. For example, some AGN exhibit broad optical/UV emission lines, some have only narrow lines, and some have no emission lines at all. In some AGN the soft X-ray emission is attenuated, but in other AGN there is an excess at soft X-ray energies relative to the "typical" power-law spectral shape. Powerful radio emission is seen in a fairly small fraction of X-ray selected AGN. The "Unified Scheme" of AGN has been formulated (Antonucci, 1993) in order to explain these observations. A cartoon illustrating the key ingredients of this model is shown in figure 1.1. The unified scheme proposes that all AGN follow a generic layout, but that the orientation of the system with respect to the observer determines the "flavour" of AGN that is seen, i.e. a "type-1" or "type-2" AGN. I now describe some of the characteristic multi-wavelength spectral features observed from AGN, the origins of this emission, and some AGN classes.

1.2.1 The central engine and X-ray emission from AGN

AGN are thought to be powered by accretion of matter onto a central super-massive black hole (SMBH), having a mass from $\sim 10^5 M_\odot$ up to as much as a few times $10^{10}\ M_\odot$ (where $M_\odot$ is the mass of the Sun). Various mechanisms cause gas and dust in the AGN host galaxy to move toward the central regions. This in-fall could, for example, be related to large scale barred/or spiral structures, or alternatively could be triggered by galaxy-galaxy merger/interaction events (Heckman, 1986; Martini et al., 2003, e.g.). In general, the infalling material which reaches the centre-most regions of the AGN will retain considerable angular momentum and so does not fall directly into the SMBH. Rather, the material first forms into a thin "accretion disk" in orbit about the black hole. Turbulence, viscosity and shear forces heat the material in the disk whilst removing its orbital velocity. The accreted material

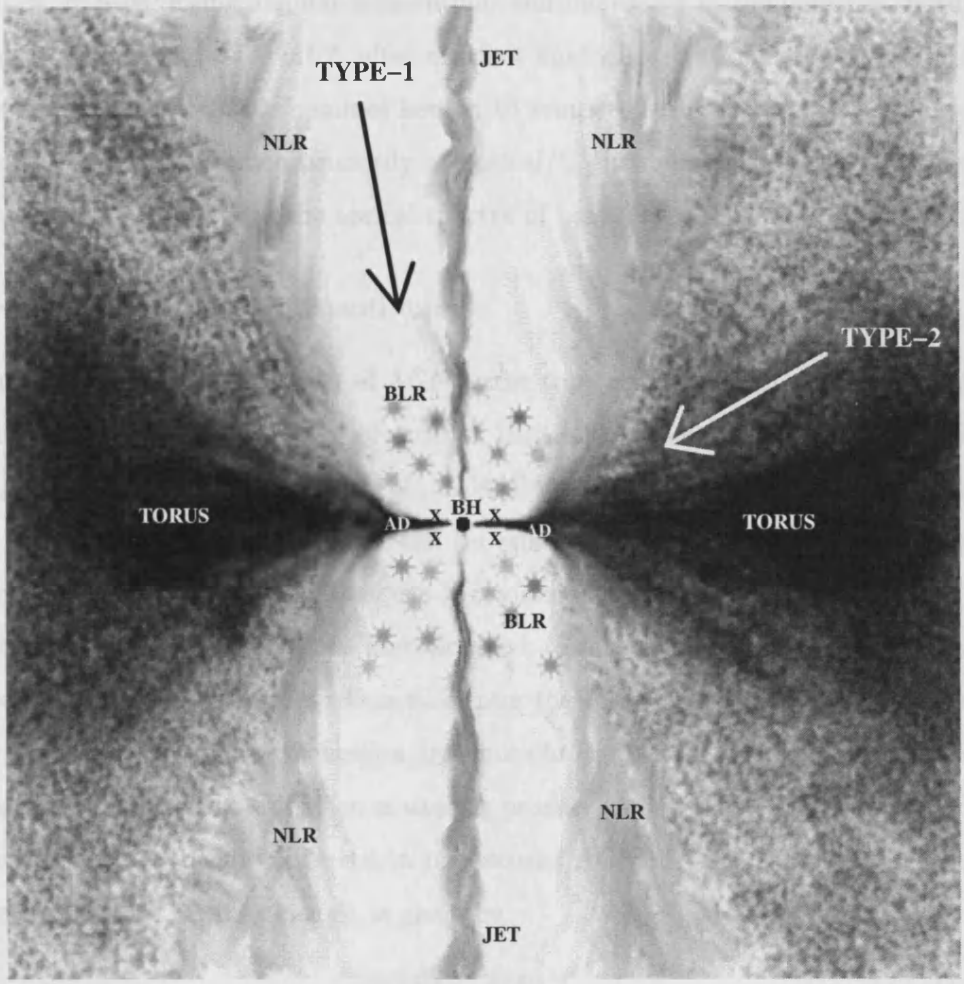

Figure 1.1: A cartoon schematic of the component parts of an AGN. The points of view of two observers are illustrated, one of which sees the cartoon AGN as a type-1 object, and one which sees it as a type-2 object. Observers viewing the AGN from the top or bottom can see the innermost regions directly (type-1). The torus blocks the view to the black hole (BH), accretion disk (AD) and broad line regions (BLR) for most of the sky (type-2). The inner edge of the torus is ionised and driven outwards by UV/X-ray photons from the accretion disk (AD), and its X-ray emitting corona (X). Narrow emission lines are thought to originate in this material (marked as NLR). The innermost parts of two relativistic jets are shown, emanating from near the black hole.

spirals inwards, losing angular momentum, and increasing in temperature, until it reaches the lowest stable orbit, after which it finally accretes onto the SMBH. The gas in the accretion disk becomes heated to temperatures of order 10^5 K, and so emits blackbody radiation primarily at optical/UV wavelengths. This is seen as a strong blue continuum in the optical spectra of most "type-1" AGN.

Production of the X-ray continuum

An almost universal signature of AGN is the copious production of X-rays in the 0.1-10 keV band, constituting of order a thirtieth of their bolometric luminosity (e.g. Elvis et al., 1994; Marconi et al., 2004; Barger et al., 2005). The X-ray spectra of AGN are typically harder than can be explained by thermal emission from the accretion disk. It is thought that the X-ray emission is produced in a "corona" of hot electrons directly above the accretion disk (Krolik, 1999). Thermal optical/UV photons produced by the accretion disk enter the corona and are then up-scattered to X-ray and Gamma-ray energies by multiple interactions with these energetic electrons. This inverse-Compton scattering process produces the "power-law" X-ray spectral shape typically observed in unobscured AGN. That is, a spectrum where the photon flux per unit energy, is given by,

$$F(E) = A(E)^{-\Gamma}, \tag{1.1}$$

where A is the normalisation at 1 keV, E is the photon energy in keV, and Γ is the photon index (typically ~ 2). At very high X-ray energies (> 100 keV), the inverse-Compton model predicts an exponential cutoff, proportional to $\exp(-E/E_{cutoff})$; an observed feature of the hard X-ray spectra of many nearby AGN.

1.2.2 X-ray absorption and scattering processes

AGN X-ray emission can penetrate through column densities of intervening material that are opaque to optical and UV photons. However, in many AGN, the amount of material between the X-ray source and the observer is sufficient for a significant

fraction of the X-ray flux to be absorbed or scattered. The degree of absorption or scattering depends on both the energy of the X-rays, as well as the quantity and state of the absorbing material.

Photoelectric absorption by neutral material

For most AGN, photoelectric absorption is the most important absorption process seen in their spectra. The probability for an X-ray of energy E to be photoelectrically absorbed by an atom in cold (neutral) material, with a Hydrogen-equivalent column density of N_H, is given by

$$P_{abs}(E) = 1 - e^{-\sigma(E)N_H}, \tag{1.2}$$

where $\sigma(E)$ is the photoelectric cross-section, which is dependent on the photon energy. The observed pattern of absorption is actually the product of many separate absorption "edges". Each edge is deepest at an energy corresponding to the ionisation potential of a particular electron shell in an element. High atomic number elements are more tightly bound, and so are most opaque at harder photon energies. Therefore the effective photoelectric cross-section is dependent on the relative elemental abundances in the absorbing material (Balucinska-Church & McCammon, 1992). Hydrogen-equivalent column densities of a few times 10^{19} cm^{-2} are required for photoelectric absorption to have a noticeable effect on the X-ray spectra at energies above 0.2 keV, the lower energy limit probed by most recent and current X-ray missions. At harder photon energies, much higher column densities are needed for a significant proportion of the flux to be absorbed. In figure 1.2, I show a plot of the X-ray spectrum expected from AGN having intrinsic power-law spectra attenuated by a range of different column densities of absorbing material. If the AGN (and the absorber) is at high redshift, then the observed column density appears lower than the true value. This is because the photons observed at energy E are emitted in the rest frame at an energy of $E(1+z)$, where the absorber is less opaque.

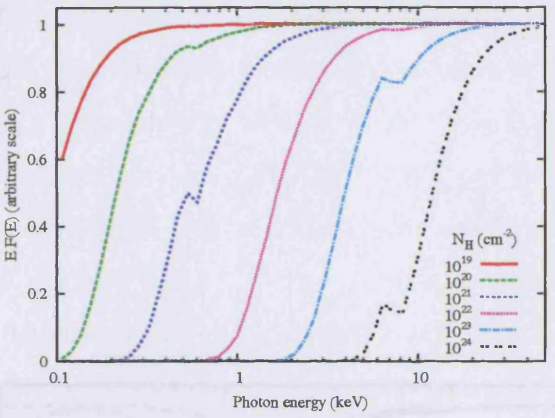

Figure 1.2: Model X-ray spectra for an AGN at $z = 0$ with an intrinsic power law of slope 2.0, absorbed by a range of column densities of neutral material (with solar abundances).

Absorption by ionised material

In many AGN, the absorbing material is partly ionised, a so-called "warm" absorber system. The pattern of absorption from an ionised absorber is dependent on the column density, the degree of ionisation, and the elemental abundances. In real absorbing systems, it is likely that some regions of the absorbing material will be more ionised than others. The various elements which make up the absorber are only partly ionised, so there remain particular energies which are preferentially absorbed because they correspond to the remaining electron binding energies. Therefore warm absorber systems are sometimes transparent to the lowest and hardest energy X-rays, but opaque to medium energy photons. The net result is a "curved" spectrum, with a deficit of photons at intermediate energies (see figure 1.3).

Compton scattering in neutral material

Compton scattering of X-ray photons by electrons becomes important for very high column densities ($N_H > 10^{24}$ cm^{-2}) of obscuring material. The electrons in the cold absorbing material typically have kinetic energies much lower than the X-ray photon energies, and so the X-rays are scattered to lower energies. In the $E_\gamma << m_e c^2$

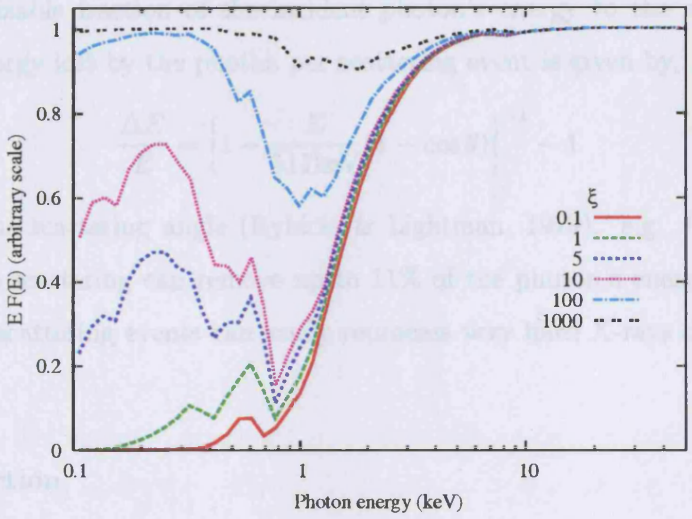

Figure 1.3: Model X-ray spectra for an AGN at $z = 0$ with an intrinsic power law of slope 2.0, absorbed by 10^{22} cm^{-2} of material with a range of ionisation parameters ξ. The absorber here is set to have solar abundances, and a temperature of 10^5K. Note the trough in the transmitted flux at $\sim$1 keV, causing the spectrum to appear "curved".

regime, the probability for a photon to be scattered follows the form of equation 1.2, with σ_E replaced with the Klein–Nishina cross-section (σ_{KN}). σ_{KN} becomes smaller with increasing photon energy, at $E_\gamma < 10$ keV, σ_{KN} is approximately equal to the Thompson cross section ($\sigma_T = 6.65 \times 10^{-25}$ cm^2), but at $E_\gamma = 200$ keV, σ_{KN} is reduced to $0.61\sigma_T$ (Rybicki & Lightman, 1979). For example, at < 10 keV, with a column density of 1.5×10^{24} cm^{-2}, $1 - e^{-1} = 63\%$ of incident photons will be Compton scattered at least once. At $E > 10$keV, each Compton scattering event transfers a sizable fraction of the incident photon's energy to the electron. The fractional energy lost by the photon per scattering event is given by,

$$\frac{\Delta E}{E} = \left[1 - \frac{E}{511\text{keV}}(1 - \cos\theta)\right]^{-1} - 1 \tag{1.3}$$

where θ is the scattering angle (Rybicki & Lightman, 1979). e.g. for 50 keV X-rays, a single scattering can remove up to 11% of the photon's energy. A handful of Compton scattering events can easily reprocess very hard X-rays down to softer energies.

X-ray reflection

Optically thick material surrounding the X-ray emitting regions can act as a "reflector", scattering a fraction of the X-rays incident on it into the line of sight of the observer (e.g. Krolik, Madau & Zycki, 1994; Magdziarz & Zdziarski, 1995). The reflecting material could be the accretion disk itself, or the inner face of the obscuring torus. The probability that an X-ray will be reflected rather than absorbed by the material is dependent on the photon energy. For a body composed of neutral material, soft X-rays are far more likely to be photoelectrically absorbed than to be reflected out of the material. Very high energy X-rays ($E > 100$ keV) will lose a large fraction of their energy through one or more Compton scattering interactions before eventually re-emerging out of the medium. Therefore, if the X-ray radiation incident on the neutral material is of the typical AGN power-law form, the reflected radiation will be most intense in the 5–50 keV range and will have a peak at around

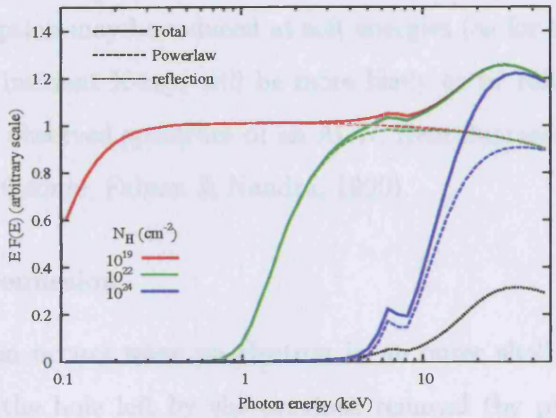

Figure 1.4: The absorbed power-law plus unabsorbed reflection component spectral model. Spectra are shown for a model AGN lying at $z = 0$, having a primary power law component with slope 2.0, exponentially cut off at an energy of 400 keV (dashed lines), and absorbed by three different column densities of neutral material. The reflection component (dotted line) is calculated for neutral reflecting material, inclined at 30°to the observer, and covering π steradians of the sky "seen" by the X-ray source. The absorber and reflecting material are set to have solar abundances. Note the hardening effect of the reflection component on the spectral shape, and the increasing importance of the reflected flux when the primary power law becomes heavily absorbed.

30 keV (Krolik, 1999). The net result of a reflection component is to harden the AGN X-ray spectrum at rest frame energies > 5 keV (see figure 1.4). In some orientations of the AGN system, the reflecting material could offer an indirect line of sight to the central regions; the reflected X-rays can bypass Compton thick material which obscures direct observation of the nucleus. If the reflector is an accretion disk, then the skin of the reflecting material will be at least partly ionised. In this case, photoelectric absorption may be reduced at soft energies (as for the warm absorber, see above), and so incident X-rays will be more likely to be reflected. If reflection is important in the observed spectrum of an AGN, then fluorescent line emission is also expected (e.g. George, Fabian & Nandra, 1990).

Fluorescent line emission

Fluorescent emission occurs when an electron in an outer shell of an atom or ion drops down to fill the hole left by the previous removal (by photoelectric ionisation) of an inner shell electron. The detection of fluorescent emission lines from AGN implies the presence of intensely irradiated, but not fully ionised material. Iron is typically the only heavy element present in appreciable quantities in AGN environments. Therefore the Fe Kα line at 6.4 keV is by far the most prominent fluorescence line in the 0.1–10 keV energy range probed by current X-ray imaging spectrometers (Krolik, 1999). Such iron lines have been reported in the X-ray spectra of many AGN (e.g. Pounds et al., 1989; Williams et al., 1992), but it is not clear whether they are an ubiquitous feature of AGN. Both narrow as well as relativistically broadened Fe K lines have been observed in AGN, implying that these lines can be emitted from material located at a range of distances from the SMBH. However, the equivalent widths of the lines detected in AGN X-ray spectral surveys are generally small, less than a few hundred eV (Braito et al., 2005; Mateos et al., 2005a; Severgnini et al., 2006), and so they do not contribute greatly to the AGN broad-band spectral energy distributions.

1.2.3 Optical classes of AGN, and their spectral properties

In unification schemes, the differences in the observed properties of AGN are attributed to the relative orientation of the viewer to the central regions. Broadly speaking, we can divide AGN into two classes: "type-1" objects and "type-2" objects see figure 1.1. In "type-1" AGN, the observer has an unobstructed view of the innermost regions of the AGN. The UV and X-ray emission from the accretion disk can be seen directly, as can the broad line and narrow line regions. However, in "type-2" AGN the "torus" of material obstructs the direct line of sight to the innermost regions, but the observer can still see the extended narrow line regions, and hard X-rays that have penetrated the obscuring material. This unified scheme predicts that if we were able to view a "type-2" AGN from closer to the "pole", we would expect to see it as a "type-1" object.

Observational limitations have an impact on the quality of the spectra that can be obtained by the current generation of ground based telescopes. Seeing-limited spectroscopic observations of AGN (at all but the lowest redshifts) measure the combined output of an AGN plus its host galaxy. For the most luminous type-1 QSOs, the AGN outshines the host by several orders of magnitude. Whereas, for some heavily obscured and low luminosity AGN, the observed optical spectrum is almost entirely due to starlight from the host galaxy (Page et al., 2003; Severgnini et al., 2003).

Broad emission line AGN

Broad line AGN (BLAGN) contain highly broadened strong emission lines in their optical/UV spectra. A common threshold for a line to be considered "broad" is if its full width at half maximum (FWHM) is ≥ 1000 km s^{-1}, (although some studies prefer FWHM ≥ 2000 km s^{-1}). The widths of the broad lines imply an origin close to the central engine of the AGN, i.e. the broadening is due to the motion of the emitting material about the SMBH. For example, some common broadened

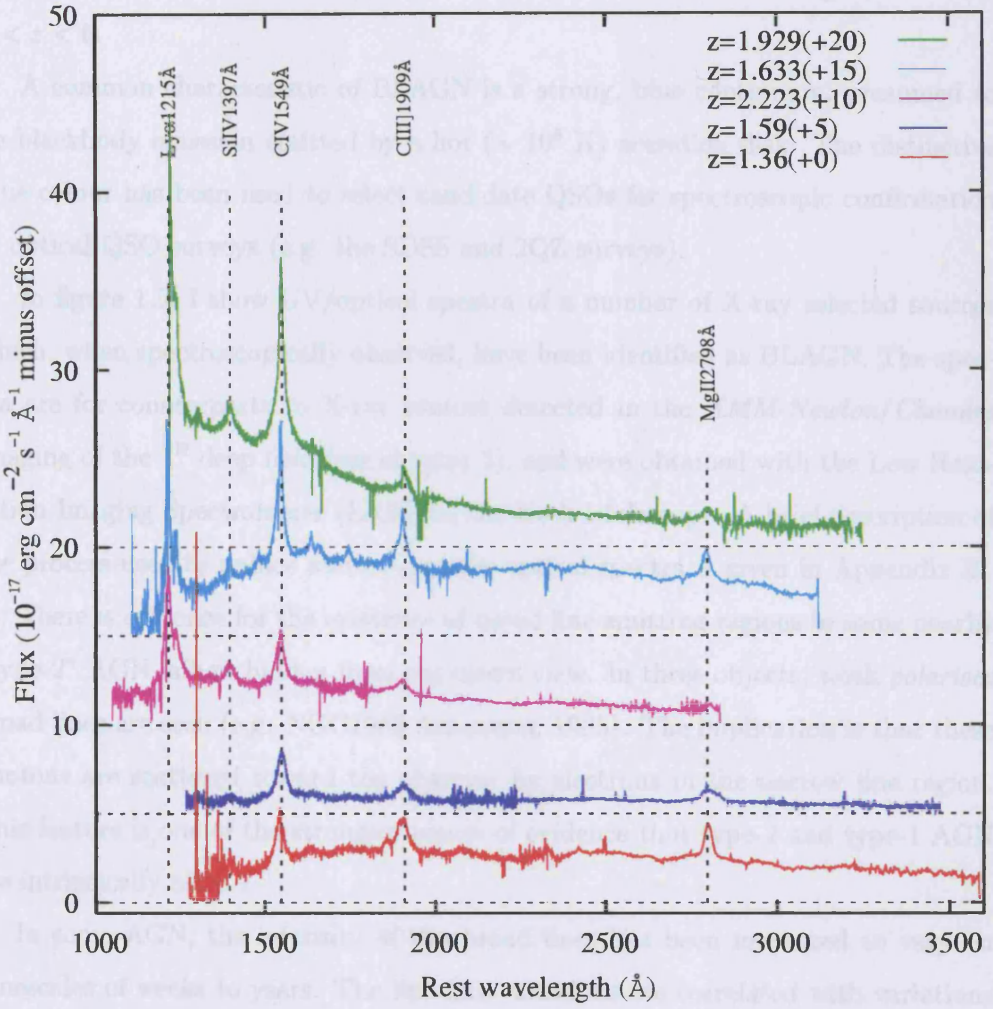

Figure 1.5: Optical spectra of a sample of five BLAGN found in the 1^H field, obtained with the Keck-LRIS instrument. Spectra are shown in the AGN rest frame, and are scaled and offset vertically for clarity. The horizontal dashed lines mark the respective shifted zero levels. The wavelengths of some commonly observed broad emission lines are indicated.

species seen in BLAGN are Lyman α, CIV, CIII], MgII, Hβ, and Hα. These lines cover a wide range of rest frame wavelengths, meaning that one or more are almost guaranteed to appear in the observed optical band for a wide range of redshifts $0 < z < 6$.

A common characteristic of BLAGN is a strong, blue continuum, presumed to be blackbody emission emitted by a hot ($\sim 10^5$ K) accretion disk. The distinctive blue colour has been used to select candidate QSOs for spectroscopic confirmation in optical QSO surveys (e.g. the SDSS and 2QZ surveys).

In figure 1.5, I show UV/optical spectra of a number of X-ray selected sources which, when spectroscopically observed, have been identified as BLAGN. The spectra are for counterparts to X-ray sources detected in the *XMM-Newton/ Chandra* imaging of the 1^{H} deep field (see chapter 5), and were obtained with the Low Resolution Imaging Spectrometer (LRIS) on the Keck-I telescope. A brief description of the process used to reduce and extract the optical spectra is given in Appendix B.

There is evidence for the existence of broad line emitting regions in some nearby "type-2" AGN, albeit hidden from our direct view. In these objects, weak *polarised* broad lines are seen (e.g. NGC1068 Antonucci, 1985). The implication is that these photons are scattered toward the observer by electrons in the narrow line region. This feature is one of the strongest pieces of evidence that type-2 and type-1 AGN are intrinsically alike.

In some AGN, the intensity of the broad lines has been measured to vary on timescales of weeks to years. The line flux variations are correlated with variations in amplitude of the UV continuum, but delayed by days to weeks . The phase delay allows an estimate of the distance of the broad line regions from the SMBH, and in combination with the velocity broadening, an estimate of the mass of the black hole (e.g. Blandford & McKee, 1982; Krolik et al., 1991; Wandel, Peterson & Malkan, 1999).

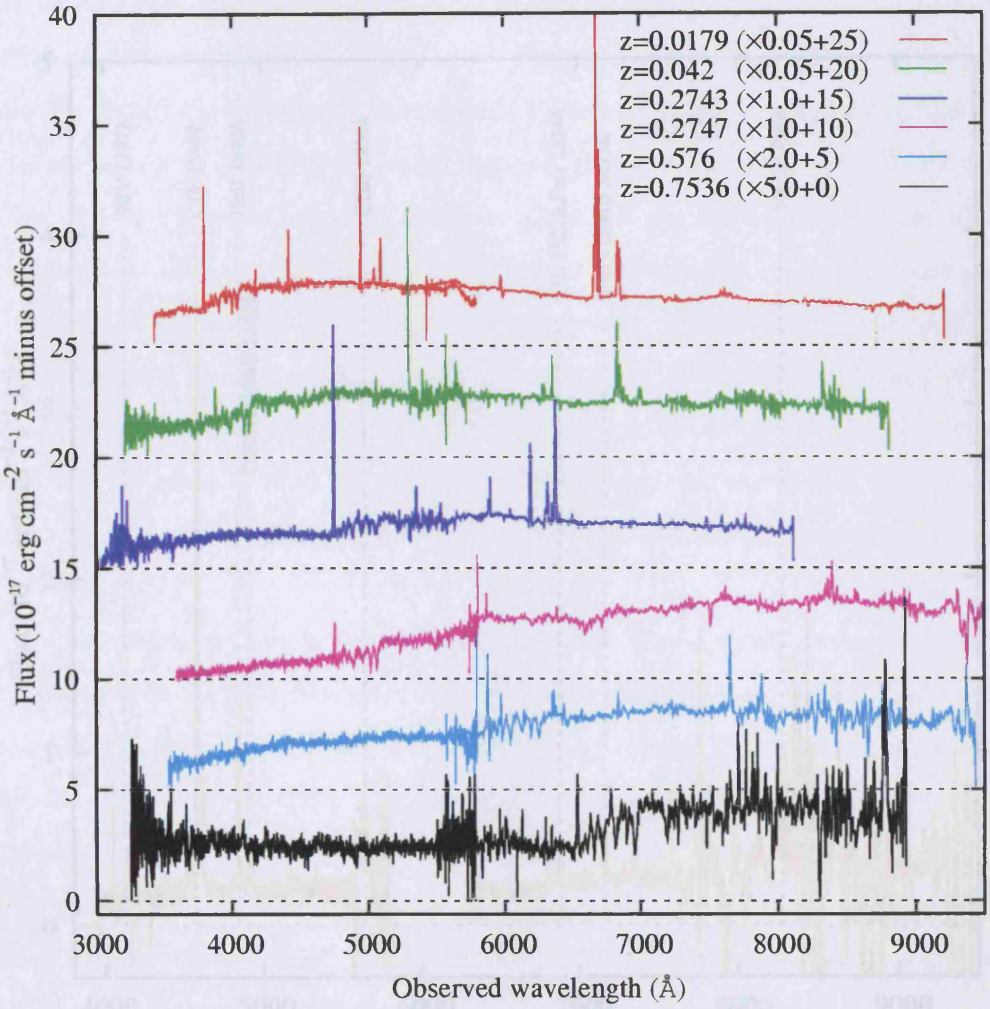

Figure 1.6: Optical spectra of a sample of six NELGs found in the 1^H deep field, obtained with the Keck-LRIS instrument. Spectra are shown in the observed frame, and are scaled/offset in flux for clarity. Horizontal dashed lines show the respective offset zero-points. Prominent narrow emission lines are [0II](3727Å), Hβ(4861Å), 0III(4959+5007Å), and Hα(6563Å). The CaII (3935+3970Å) absorption system is also evident in several of these spectra.

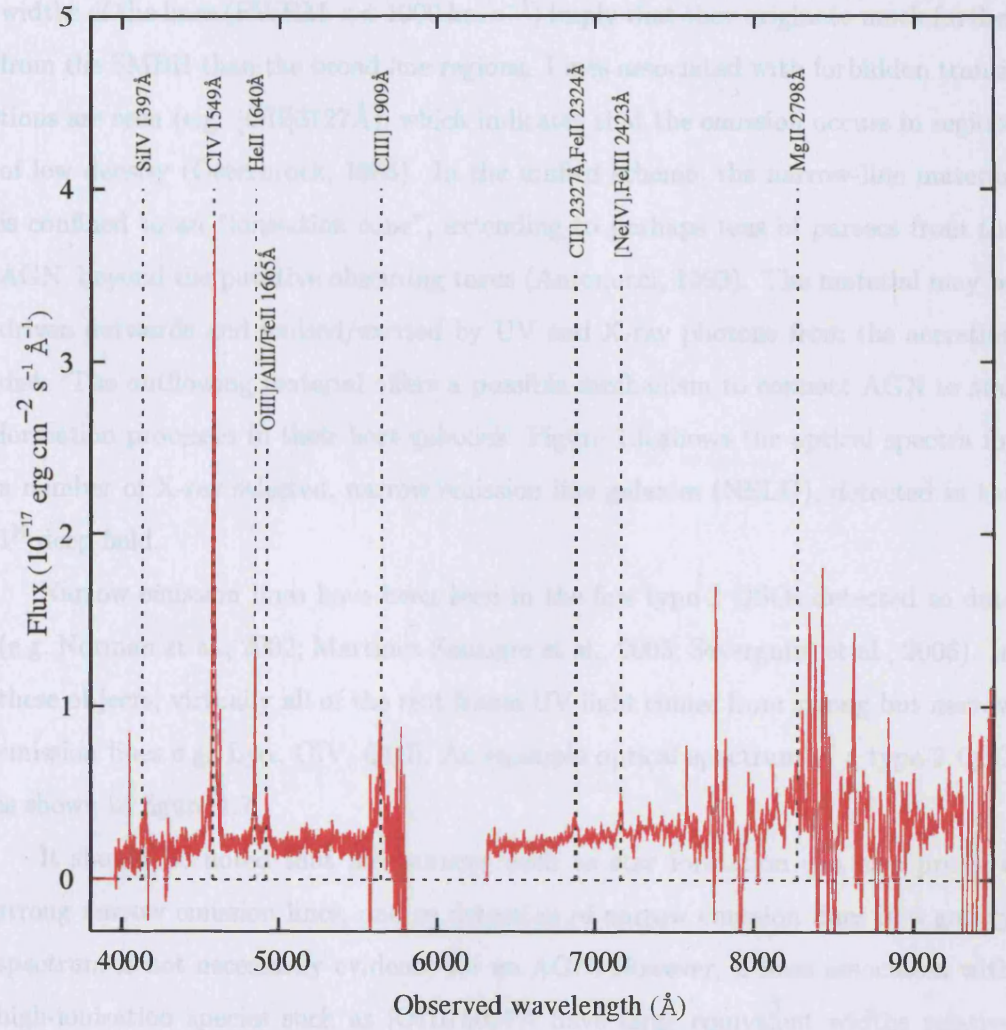

Figure 1.7: Optical spectra of a X-ray selected, high redshift (z=1.956) type-2 QSO found in the 1^{H} deep field. The spectrum is shown in the observed frame, with prominent emission lines marked. Some of the lines do show signs of slight broadening, but with FWHM $\leq$ 500 km s^{-1}.

Narrow emission line galaxies

Narrow emission lines are observed in the optical spectra of most AGN. The small widths of the lines (FWHM $<<$ 1000 km s^{-1}) imply that they originate much further from the SMBH than the broad line regions. Lines associated with forbidden transitions are seen (e.g. [OII]3727Å), which indicates that the emission occurs in regions of low density (Osterbrock, 1988). In the unified scheme, the narrow-line material is confined to an "ionisation cone", extending to perhaps tens of parsecs from the AGN, beyond the putative obscuring torus (Antonucci, 1993). The material may be driven outwards and ionised/excited by UV and X-ray photons from the accretion disk. The outflowing material offers a possible mechanism to connect AGN to star formation processes in their host galaxies. Figure 1.6 shows the optical spectra for a number of X-ray selected, narrow emission line galaxies (NELG), detected in the 1^{H} deep field.

Narrow emission lines have been seen in the few type-2 QSOs detected to date (e.g. Norman et al., 2002; Martinez-Sansigre et al., 2005; Severgnini et al., 2006). In these objects, virtually all of the rest frame UV light comes from strong but narrow emission lines e.g. Lyα, CIV, CIII]. An example optical spectrum of a type-2 QSO is shown in figure 1.7.

It should be noted that mechanisms such as star formation can also produce strong narrow emission lines, and so detection of narrow emission lines in a galaxy spectrum is not necessarily evidence for an AGN. However, if lines associated with high-ionisation species such as [OIII] 5007Å have large equivalent widths relative to the lower ionisation lines (e.g Hβ), then the presence of an AGN is much more certain (e.g. Krolik, 1999).

1.2.4 Broad absorption line quasars

Broad absorption line quasars (BALQSOs) exhibit the same broad emission lines as seen in BLAGN. However, as the name suggests, they also have deep, broadened

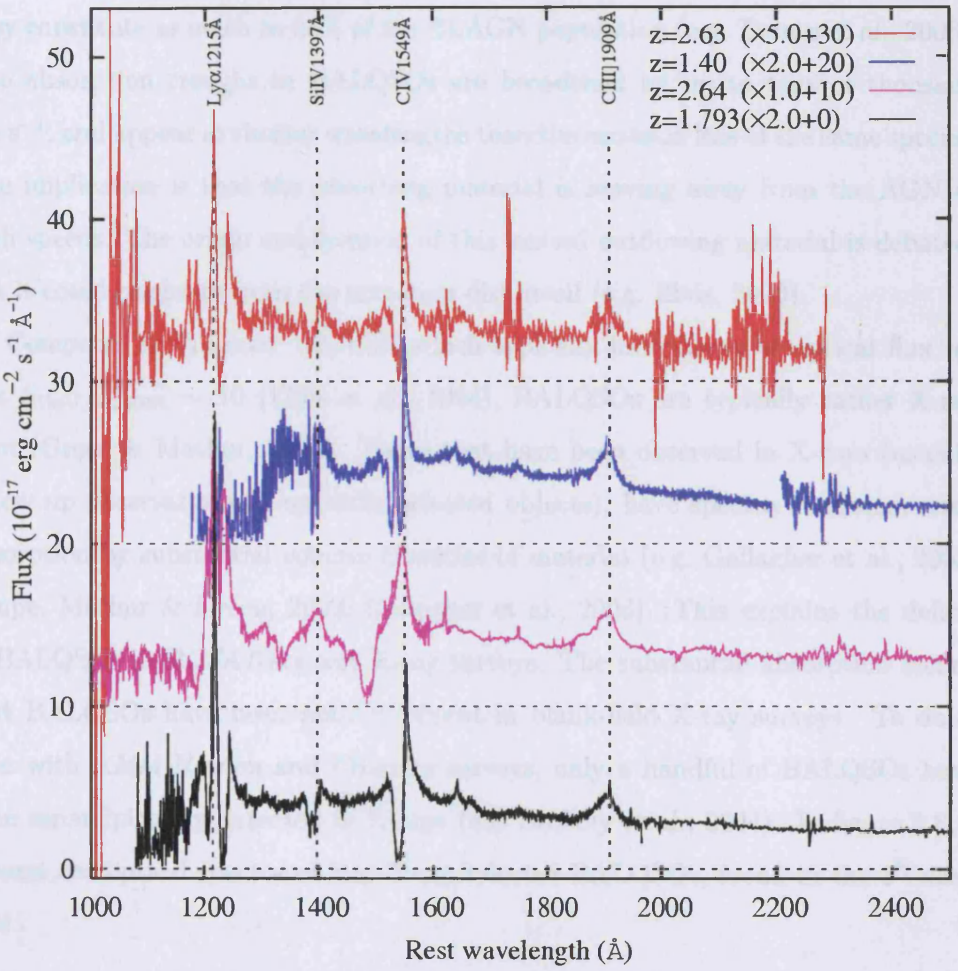

Figure 1.8: Optical spectra of the four BALQSOs found in the 1^{H} field. Three of the spectra were obtained with the Keck-LRIS instrument, and one with the WHT-WYFFOS/AF2 fibre spectrograph. The spectra are displayed in the rest frame of these QSOs and are scaled and shifted vertically for clarity. Horizontal dashed lines show the respective offset zero-points. These objects show blue-shifted broad absorption associated with the CIV 1549Å and SiIV 1397Å lines.

absorption troughs in their continua, associated with the CIV and SiIV transitions for example. BALQSOs are rather common in optically selected quasar samples; they constitute as much as 25% of the BLAGN population (e.g. Trump et al., 2006). The absorption troughs in BALQSOs are broadened by up to tens of thousand km s^{-1}, and appear at shorter wavelengths than the emission line of the same species. The implication is that the absorbing material is moving away from the AGN at high speeds. The origin and location of this ionised outflowing material is debated, but it could originate from the accretion disk itself (e.g. Elvis, 2000).

Compared to "normal" BLAGN which typically have X-ray to optical flux ratios $S_{1\mathrm{keV}}/S_{2500\mathrm{A}} \sim 10$ (Elvis et al., 1994), BALQSOs are typically rather X-ray faint (Green & Mathur, 1996). Those that have been observed in X-rays (usually follow up observations of optically selected objects), have spectra consistent with absorption by substantial column densities of material (e.g. Gallagher et al., 2002; Grupe, Mathur & Green, 2003; Gallagher et al., 2006). This explains the deficit of BALQSOs in *ROSAT*-era soft X-ray surveys. The substantial absorption means that BALQSOs have been notably absent in blank-field X-ray surveys. To date, even with *XMM-Newton* and *Chandra* surveys, only a handful of BALQSOs have been serendipitously detected in X-rays (e.g. Szokoly et al., 2004). In figure 1.8, I present the optical spectra of four X-ray selected BALQSOs, found in the 1^{H} deep field.

Absorption and emission from the torus

Surrounding the central parts of AGN is thought to be a roughly axisymmetric parsec scale structure of gas and dust. A toroidal shape is usually proposed because it is axisymmetric and easily parametrised. The size, shape and location of the obscuring material remains unclear, but several constraints can be made.

For example, the intense UV/X-ray emission from the central engine coupled with the sublimation temperature of dust ($\sim 1000 - 1500$ K dependent on dust composition Granato 1997), effectively defines the innermost possible edge for a

dusty torus. Dust closer in will simply evaporate. The radius of the inner edge is thus expected to be proportional to the square root of the luminosity of the central engine. For a $L_X = 10^{44}$ erg s^{-1} AGN the innermost possible radius would be ~ 1 pc (Krolik, 1999). However, the inner edge of the torus could extend closer to the SMBH if it is shielded somehow from direct radiation coming from the accretion disk. It is less clear how far outwards such tori extend. High resolution Hubble Space Telescope (HST) imaging of the innermost regions of nearby Seyferts has directly revealed the presence of dusty toroidal structures up to a few tens of parsecs in size (e.g. Ferrarese, Ford & Jaffe, 1996). However, this approach is limited to the nearest AGN, all of which are of fairly low luminosity; it is unknown how the size/shape of an "average" torus might scale with the size/luminosity of the central black hole/accretion disk. An upper limit can be made from the total mass of the obscuring material in the torus. If there is much more mass in the torus than the black hole, then the system will become gravitationally unstable (e.g. Risaliti, Maiolino & Salvati, 1999).

The torus is heated by the incident UV/X-ray radiation from the AGN. This is then re-radiated as thermal emission at longer wavelengths (e.g. $\lambda_{peak} \sim 100\mu$m @ $T \sim 50$ K). Observations of AGN in the mid- to far-infrared are most sensitive to this warm dust. Unfortunately this is a very difficult part of the spectrum to observe from the ground. Space based MIR/FIR observatories, such as ISO and *Spitzer* are limited by relatively small mirrors (diameter $<$ 1m), and so cannot resolve these dusty structures, even in the nearest AGN. However, better constraints can be made by observing the absorption/emission of the torus at several wavelengths. For example, the intrinsic X-ray luminosity of the AGN and the column density through the torus can be calculated from the X-ray spectrum. MIR/FIR/sub-mm photometry can reveal the total re-radiated blackbody output from the torus, which is related to its temperature and projected size. There is now direct observational evidence that "torus" structures do exist in type-1 objects. MIR silicate emission features have recently been measured by *Spitzer*-IRS in a number of type-1 objects (Yan et al., 2005; Sturm et al., 2005), and are thought to originate from the inner

edge of a torus heated by the central engine.

The torus may provide a vital part in the fuelling process of AGN; providing a staging post and/or a reservoir in the process which moves the gas and dust from the disk of a galaxy, which has considerable angular momentum, to the AGN. The inner edge of the torus, which is perhaps 1pc from the SMBH, may supply material directly to the outer parts of the accretion disk. If tori do provide an efficient fuelling route for AGN, then we might expect a large fraction of cosmic accretion power to occur in obscured environments.

AGN which do not follow the type-1/type-2 paradigm

Various joint X-ray/optical surveys (e.g. Page et al., 1997; Hasinger et al., 1998; Branduardi-Raymont et al., 1994), have found that the vast majority of BLAGN are unobscured in X-rays. This is in line with the Unified scheme, in which the X-rays, blue continuum, and broad lines are produced relatively close to the SMBH compared to the size of the dusty torus. If an observer has an unobstructed view of any one of these emission sites, then it might be expected that the others will also be visible. However, numerous exceptions to this paradigm have been charted (e.g. Page, Mittaz & Carrera, 2001; Comastri et al., 2002; Page et al., 2006). Most notably, some optically selected BLAGN show significant absorption in their X-ray spectra, and some unabsorbed X-ray selected AGN do not have broad lines evident in their optical spectra (X-ray bright, optically normal galaxies, XBONGs). Several explanations have been offered for these apparent contradictions. For an X-ray absorbed, optical type-1 object, it could be that, by chance, we are viewing it at a particular orientation: the centre-most X-ray regions are obscured by the torus, but the slightly further out BL region is still viewable. However, this picture seems rather contrived, and whilst feasible for a few such objects, cannot be offered as a reasonable explanation for the many X-ray absorbed, but optical type-1 objects that have been observed. Alternatively, the X-ray obscuring material in these objects could have a different mix of gas and dust to the interstellar medium in local galaxies. For

example, an obscuring medium which is primarily made of gas, and contains very little dust, will be opaque to soft X-rays, but relatively transparent to optical/UV photons. Another explanation is AGN variability: AGN have been observed to undergo considerable brightness and spectral variations on timescales of weeks to months (e.g. Risaliti et al., 2005). However, it is typical for X-ray and optical observations of faint AGN to be carried out at different epochs, and so may be measuring the AGN whilst it is in very different spectral or luminosity states.

A (somewhat speculative) possibility is that some of these unusual sources are galaxies containing more than one active nucleus, the result, for example, of a galaxy merger. In general, these two (or more) AGN will have different luminosities and SEDs. So for example, a powerful obscured AGN could be responsible for the X-ray spectrum, whilst a less luminous, but unobscured AGN could be producing the broad lines. The number of such systems that we expect to observe is dependent on the frequency of AGN-AGN mergers, and the timescales for such systems to coalesce into a single SMBH. In a "hierarchical" Universe, in which massive galaxies and SMBHs have grown through cannibalism, then one might expect such multiple-AGN systems to be commonplace. However, in the local Universe, where we are able to spatially resolve such systems (with current instrumentation), there have been only a few examples detected (e.g. NGC6240, Risaliti et al. 2006). In the epoch when mergers were most frequent, $z > 2$, an angular resolution of better than $\sim 0.1''$ is required to resolve two nuclei separated by 1 kpc, far beyond the capabilities of the current or even next generation of X-ray instrumentation.

1.3 Charting the AGN population

There is a range of observational measurements that can be used to constrain the demographics of the AGN population over cosmological timescales. Some constraints are more powerful than others, but all have their limitations.

Perhaps the least selective and most efficient method for the detection of AGN

of all types, and at high redshifts, are deep X-ray observations of "blank" high Galactic latitude fields. AGN can be selected very efficiently in such fields because they contain few Galactic X-ray sources, and most field galaxies are X-ray quiet. The X-rays we see from AGN originate in the centre-most regions, and so provide perhaps the most direct probe of the accretion process, and the environment near the SMBH.

1.3.1 Extragalactic X-ray surveys

The field of extragalactic X-ray astronomy was born when Giacconi et al. (1962) reported the existence of a diffuse cosmic X-ray background (XRB). Subsequent X-ray observations were carried out using either rocket- or balloon-borne instruments and were capable of detecting only the brightest objects in the X-ray sky, primarily sources in our own Galaxy. The first orbiting X-ray observatories (e.g. *UHURU*, *Ariel-V*, and *HEAO-1*) were launched in the 1970s. The combination of longer available exposure times and greater instrument sensitivity allowed astronomers to see deeper into the X-ray sky. The majority of the sources detected with these observatories were binary systems located in our own Galaxy. A great leap forward in *extragalactic* X-ray astronomy came with the launch in 1978 of *Einstein*, the first X-ray satellite to incorporate focusing mirror optics. The combination of a much larger large effective area and better spatial resolution than previous missions (100 cm^2 and FWHM $\sim 1'$ respectively for the Imaging Proportional Counter), meant that observations with *Einstein* detected thousands of extragalactic X-ray sources. It was postulated that the XRB, rather than being truly diffuse, could actually be the integrated emission from numerous faint extragalactic sources. *ROSAT*, launched in 1990, had a larger effective area, further improved spatial resolution and a wide field of view (240 cm^2, FWHM$\sim 25''$, and 2° diameter respectively for the Position Sensitive Proportional Counter), giving it an even better "grasp" as a survey telescope. A number of surveys were carried out with *ROSAT*, ranging from a shallow map

of the entire sky, to very deep pointed observations of small areas. In the deepest surveys, around 70% of the 0.1–2.5 keV XRB intensity was resolved into point-like sources (Hasinger et al., 1998; Branduardi-Raymont et al., 1994). Due to technical constraints, X-ray astronomy at > 2 keV has lagged behind that at softer X-ray energies. *ASCA* and *BeppoSax* were the first observatories to explore the 2–10 keV energy band to reasonable depths. The intrinsic energy resolution of the detectors on these observatories meant that the hard X-ray spectral shapes of many AGN could be measured at moderate resolution. The added dimension of X-ray spectral information is a crucial key to understanding the nature of the X-ray sources.

The *XMM-Newton* and *Chandra* observatories

The NASA *Chandra* and ESA X-ray Multi Mirror Newton (*XMM-Newton*) X-ray satellites were both launched in 1999. *XMM-Newton* has the largest effective mirror area (~ 4500 cm^2 at 1 keV) ever flown on an X-ray focusing observatory, whereas *Chandra* has the best ever spatial resolution ($< 1''$ half energy width). This thesis makes great use of data collected with the *XMM-Newton* observatory. In this section I give an overview of the *XMM-Newton* observatory and its detectors. In chapter 2.1 I describe in detail the process I have used to reduce *XMM-Newton* X-ray imaging data. I also make some use of the results of surveys carried out with the *Chandra* observatory, and so I briefly summarise its salient features later in this section.

XMM-Newton (Jansen et al., 2001) consists of three co-aligned X-ray telescopes, and an optical/UV monitor. At the focus of each X-ray telescope lies a European Photon Imaging Camera (EPIC), with useful sensitivity to photons in the energy range 0.2–10 keV, and having a field of view (FOV) of $\sim 30'$ diameter. Two of the EPICs use arrays of front-illuminated CCDs of the conventional Metal Oxide Semiconductor (MOS) type (Turner et al., 2001), the third uses a novel, back-illuminated *pn* CCD array (Struder et al., 2001). I refer to these as the MOS1, MOS2 and *pn* detectors respectively. For the two MOS detectors, approximately half of the incoming light from the mirrors is reflected by gratings which feed high

resolution spectrometers (den Herder et al., 2001). But unfortunately, the latter are useful only for the study of sources far brighter than those studied here. The intrinsic energy discrimination of the EPIC CCDs allows moderate resolution (R$\sim 15-40$) X-ray spectroscopy. *XMM-Newton* has a highly elliptical orbit (period $\sim$ 2 days), which permits long (up to 140 ks) continuous observations at high altitudes, above the Earth's radiation belts. However, *XMM-Newton* does suffer from periods of high particle background which reduces (by up to $\sim$ 30%) the effective time available for scientific observations of faint sources. The EPIC detectors have filters of three thicknesses used to block optical light from entering the X-ray cameras. If left unblocked, the optical photons would degrade the sensitivity and spectral resolution of EPIC.

The design of the *Chandra* telescope, also launched in 1999, sacrificed mirror collecting area in favour of very compact point spread function (Weisskopf et al., 2000). The FWHM of the on-axis point spread function is $< 1''$, but the effective mirror area is $\sim$ 800 cm^2 at 1 keV. Imaging with *Chandra* is carried out primarily with the ACIS-I detector (Advanced CCD Imaging Spectrometer - imaging mode, Batz et al. 1999). ACIS-I consists of a mosaic of front illuminated, conventional-type CCDs having a field of view of $17' \times 17'$. In practice ACIS-I is sensitive over a narrower energy range than EPIC, since it has an effective area, including detector quantum efficiency, of $\geq$ 100 cm^2 over only the 0.7–7 keV range. The high spatial resolution, and low background in the images collected with *Chandra* make it much more sensitive to the detection of faint point-like sources than *XMM-Newton* or previous X-ray observatories. The high positional accuracy also means that for most X-ray sources, a unique counterpart can be identified at other wavebands. Despite the much better image quality compared to *XMM-Newton*, *Chandra*'s smaller mirror area and narrower field of view mean that it is somewhat less useful for determining the X-ray spectral properties of large numbers of faint AGN.

X-ray surveys in the current era

A number of deep X-ray surveys have been made using *XMM-Newton* and *Chandra*, with the investment of many mega seconds (Ms) of observing time. The key parameters which define the sensitivity and capabilities of an X-ray survey are the telescope(s) used, the exposure times, and the sky coverage. An important consideration for the scientific usefulness of these surveys is the presence of observations at other wavelengths, for example, the acquisition of spectroscopic and/or photometric redshifts of the X-ray sources. The X-ray surveys which reach the deepest in terms of the faint flux limit are the "pencil beam" *Chandra* deep fields -North and -South (CDFN and CDFS), in which 2 Ms and 1 Ms of *Chandra* observations have been made respectively. *XMM-Newton* has also carried out a number of deep blank field surveys, of a hundred kiloseconds or more (e.g. the Lockman hole, the CDFS, the 13^{H} deep field). A number of wider-field but shallower extragalactic surveys have also been carried out using *Chandra* and/or *XMM-Newton*, e.g. the Groth strip (Nandra et al., 2005), ELAIS (Manners et al., 2003), XMM-COSMOS[1], the Subaru-*XMM-Newton* deep field[2], and the *XMM-Newton* large scale survey (Chiappetti et al., 2005). For this thesis I use data from six fields which have amongst the deepest *XMM-Newton* coverage, as described later. Deep X-ray surveys which combine data from both the *Chandra* and *XMM-Newton* observatories can take advantage of their mutually complementary capabilities; in chapter 4, I make use of the joint *XMM-Newton*/*Chandra* coverage of the CDFS field (Giacconi et al., 2002).

1.3.2 AGN surveys at other wavelengths

In general, AGN surveys are carried out in a waveband where the AGN have some distinct "signature" which sets them apart from all the non-AGN sources detected in the same waveband. In most cases, a optical spectroscopic observation is also

[1]http://www.mpe.mpg.de/XMMCosmos/

[2]http://www.naoj.org/staff/chris/SXDS/

required to confirm the nature (and measure the redshift) of the AGN candidates selected by such a survey.

AGN were first detected in observations at radio wavelengths, and deep radio surveys remain an efficient method for the detection of large numbers of AGN (e.g. Becker et al., 1995; Seymour, McHardy & Gunn, 2004). However, strong radio emission is not an ubiquitous feature of all AGN; for example, $< 10\%$ of the X-ray sources in the Cambridge-Cambridge *ROSAT* Serendipity survey were also detected in 1.4 GHz Very Large Array (VLA) observations (Ciliegi et al., 1995). What is more, high frequency radio surveys may favour the detection of sources in which the radio emission is beamed toward the observer, preferentially selecting only a subset of possible AGN orientations.

Surveys in the optical have increased dramatically in size and sensitivity in recent years. For example, the Sloan digital sky survey (SDSS), and the 2dF quasar redshift survey (2QZ), cover thousands of square degrees of the sky, and detect tens of thousands of AGN (Abazajian et al., 2003; Croom et al., 2004). Smaller area, but deeper optical surveys have also been undertaken, for example the COMBO-17 photometric redshift survey (Wolf et al., 2003). However, optical surveys are usually designed to detect broad line AGN, for example by selecting candidate AGN on the basis of their blue optical colours. Therefore, relatively few narrow line AGN appear in these optical surveys (e.g. Williams, Pogge, & Mathur, 2002; Komossa et al., 2006).

Deep surveys in the infrared have, until recently, been limited in scope because of the difficulty of observing at these wavelengths. The principal component from AGN in the mid and far infra-red (MIR/FIR) is thermal emission from warm dust in the obscuring torus. AGN stand out from other astronomical objects; dust that is heated by an intense, compact source i.e. an AGN, is hotter, and thus re-radiates at shorter wavelengths than dust heated by a more dispersed source of radiation, i.e. star formation. Therefore, candidate AGN can be selected on the basis of their MIR/FIR spectral energy distributions.

In the NIR, the continuum emission we see from AGN is thermal emission from the accretion disk. However the disk has a broad temperature range, hot in the centre, cooler further away from the SMBH. Thus, the observed emission in the NIR is actually the superposition of the Rayleigh-Jeans tails of a distribution many blackbody spectra, and can be approximated by a power law continuing far into the MIR. In contrast, in the NIR, the emission from galaxies is dominated by old stellar populations, with a peak at rest frame wavelengths of $\sim 1.5\mu$m. The relative "flatness" of the AGN spectral energy distributions can thus be used to select AGN in the NIR (e.g. Rowan-Robinson et al., 2005; Stern et al., 2005). What is more, at NIR wavelengths, emission from the central AGN regions can usually be seen directly, even when buried beneath absorbing material that is virtually opaque to optical photons. The current and forthcoming NIR surveys (e.g. UKIDSS, VISTA) for the first time cover a wide enough area with sufficient sensitivity to detect and examine large numbers of type-2 AGN.

1.4 The demographics and evolution of the AGN population

So in light of all these AGN survey techniques, what is our current understanding of the AGN population?

1.4.1 The AGN luminosity function

For samples of AGN where redshift determinations have been made, it is possible to measure the AGN "luminosity function" (LF). The LF describes the number density of AGN as a function of their luminosity (L) and redshift (z).

Optical surveys, which primarily select luminous type-1 AGN, have revealed that at low redshifts ($z \sim 0.5$), the LF is well described by a smooth double powerlaw, having a "knee" at an absolute magnitude of $M_B \sim -23.2$ (e.g. Croom et al.,

2004). At luminosities below the knee, the slope is flatter, and above this level, it is steeper. At high redshifts, the LF has a similar shape, but is shifted to much greater luminosities. For example, in the Croom et al. (2004) model, the knee luminosity increases to $M_B \sim -23.3$ at $z = 1$, and to $M_B \sim -25.6$ at $z = 2$. The QSO number density at higher redshifts is less well defined, but appears to slowly decrease toward very high redshifts. The redshift dependence of the optical QSO LF is well described by "pure luminosity evolution" (PLE) models, in which $LF(L, z) = LF(L/e(z), z = 0)$, where $e(z)$ describes the luminosity evolution, and can be parametrised as a polynomial in z, or an exponential in lookback time. Another proposed evolution model is one in which the shape of the LF remains constant, but where the space density increases towards high redshift, i.e. $LF(L, z) = LF(L, z = 0) \times e(z)$ (Pure Density Evolution, PDE). The PDE model is now disfavoured as a good description of the evolution of the optical QSO LF. This is mainly because QSO surveys have become sensitive enough to reach below the knee, even at high redshifts (e.g. Boyle, Shanks & Peterson, 1988). These surveys have revealed that at high redshifts, the knee of the LF occurs at higher luminosities than at low redshifts, a pattern impossible to reproduce with a PDE model. In summary, the optical QSO LF tells us that luminous AGN activity increased slowly from very high redshifts up to its peak at $1.5 < z < 2.5$, and has decreased rapidly ever since.

X-ray surveys typically probe a rather different part of the AGN population to the optical QSO surveys. While the optical surveys are able to cover many thousands of square degrees, they are sensitive only to the most luminous type-1 QSOs (which have a low sky density). On the other hand, deep X-ray surveys are typically small in area ($< 1\mathrm{deg}^2$), but can detect both type-1 and type-2 AGN, to far lower luminosity limits.

The AGN detected in X-ray surveys must be identified spectroscopically if we are to determine their distribution in redshift and luminosity. This is a major bottle neck for X-ray surveys; identification programs often take up many times the effort that was expended on the initial analysis of the X-ray observations. A number

of issues present themselves for identification programs. Firstly, each X-ray source must be matched to the correct optical counterpart, which requires deep optical imaging. This step has become easier with the high positional accuracy of the current X-ray telescopes, particularly *Chandra*, which typically constrains the X-ray position to better than 1″ (e.g. McHardy et al., 2003; Szokoly et al., 2004). However, many X-ray sources have very faint optical counterparts ($R > 24$); the high spatial density of optical sources at these faint magnitudes increases the chance of misidentification. In extreme cases, the X-ray source may have no counterpart to the limit of the optical imaging (Mainieri et al., 2005). Once the most likely optical counterpart has been determined, it must be identified spectroscopically. Whilst this is relatively straightforward for the optically bright ($R < 22$) AGN, identification of fainter sources requires long exposures with large telescopes. The advent of high throughput multi-slit spectrographs (e.g. Keck-DEIMOS, VLT-FORS, VLT-VIMOS) has made this process somewhat more achievable. For the X-ray sources with the faintest optical counterparts ($R > 24$), optical spectroscopic identification is effectively impossible, even with 8m class telescopes. Therefore, it is usual for any faint X-ray survey to have a fraction of unidentified sources. This "incompleteness" must be accounted for when interpreting the results of the survey. The redshift incompleteness can introduce additional selection effects into the survey because some AGN types are easier to identify at certain redshifts than others. For example, above redshift ~ 1.2, the prominent [OII],Hβ,Hα, and OIII AGN emission lines are all redshifted out of the bandpass of most optical spectrometers, hindering the identification of AGN which do not also have prominent rest frame near-ultraviolet (NUV) lines.

For the reasons above, the numbers of AGN in the samples used to measure the X-ray luminosity function (XLF) are typically small. Relative to the QSO LF (which is based on tens of thousands of objects), the AGN XLF is rather poorly determined. In the *ROSAT*-era, the largest soft X-ray selected samples used to determine the XLF contained of order 700 sources (Page et al., 1997; Miyaji et al.,

2000). The situation is even worse for AGN selected at harder X-ray energies: for example, there were only 247 sources in the 2–10 keV selected sample used by Ueda et al. (2003) to determine the hard XLF. However, these studies have shown that the X-ray selected AGN luminosity function is broadly similar to that seen in the optical QSO surveys. At low redshift, the XLF is well described by a double power law with a knee at $L_X \sim 10^{44}$ erg s^{-1}. The AGN number density declines rapidly at luminosities above the knee, and increases slowly at luminosities below the knee. As with optical surveys, there is strong evolution in the XLF, with the luminosity density increasing with redshift, up to $z > 1$.

The precise form of the XLF evolution is poorly known. Page et al. (1997) found that pure luminosity evolution was an adequate description of the XLF of soft X-ray selected BLAGN, with the evolution continuing up to a redshift of 1.5–1.8. However, their study was conducted assuming a matter dominated cosmology, now deprecated in light of the recent Wilkinson Microwave Anisotropy Probe results (Spergel et al., 2003). Miyaji et al. (2000) also took a sample of soft X-ray selected AGN (not exclusively broad line objects), and used it to calculate the XLF. These authors tested several model forms for the luminosity function, and found that a "luminosity dependent density evolution" (LDDE) scheme provided the best match to the redshift/luminosity distribution of their sample. The LDDE model is invoked to explain an apparent deficit of low to intermediate luminosity AGN at high redshifts. In the Miyaji et al. (2000) LDDE model the evolution is of the form $LF(L, z) = LF(L, z = 0) \times e(L, z)$, where the number density of luminous AGN evolves much faster than that of lower luminosity objects. Unfortunately, Miyaji et al. (2000) fitted the PLE model only for matter dominated cosmologies, and so direct comparison of the LDDE and PLE models is not possible for the $(\Omega_M, \Omega_\Lambda) =$ (0.3,0.7) cosmology.

The contribution of AGN to the XRB, per logarithmic unit luminosity, is at a maximum at the knee of the XLF. Although the number density of AGN does increase below the knee, it does so more slowly than the decrease in luminosity of

the individual AGN. Therefore, perhaps the most useful X-ray surveys for the study of the AGN population are those which contain large numbers of AGN near the knee luminosity

1.4.2 The distribution of absorption in the AGN population

Optical studies of nearby Seyfert galaxies have found that the ratio of type-2 to type-1 Seyferts is approximately 4 (Maiolino & Rieke, 1995). The hard X-ray study of Seyfert-2s by Risaliti, Maiolino & Salvati (1999) discovered that they have a distribution of absorbing columns, but with $\sim 75\%$ having $N_H > 10^{23}$ cm^{-2}. However, this study was limited to the very local universe, $\langle d \rangle = 24$ Mpc, and to low nuclear luminosities, $M_B > -22$; the behaviour in the rest of the redshift-luminosity plane is less well understood. The distribution of absorption in X-ray selected AGN is also poorly constrained, the prime difficulty being that the greater an AGN's N_H, the lower is its chance of being detected, or optically identified. Whereas soft X-ray surveys e.g. using *ROSAT*, have primarily charted the unabsorbed objects, we need to look in the hard X-rays ($E > 2$ keV) if we are to detect reasonably sized samples of absorbed AGN.

Despite these difficulties, estimates of the N_H distribution have been made using hard X-ray selected, but fairly small, AGN samples (e.g. Ueda et al., 2003; Treister et al., 2004). These studies have typically determined N_H from the examination of X-ray hardness ratios as a proxy for full spectral fitting. Conflicting results have been arrived at. For example, Ueda et al. (2003) describe the distribution of absorption within their sample with a luminosity dependent N_H model, in which high intrinsic-luminosity AGN are less likely to be heavily absorbed; however the model requires some additional Compton thick AGN to reproduce fully the XRB when extrapolated to harder energies. In contrast, Treister et al. (2004) find that a luminosity-independent N_H model, in which luminous objects are just as likely to be absorbed as lower luminosity AGN, provides an adequate match to their very

faint AGN sample.

So, despite the progress made in resolving, and to some extent optically identifying, the hard X-ray population, it has still not been possible to definitively delimit the true distribution of absorption in the AGN population. The problem is particularly acute for heavily absorbed AGN at high-z, very few of which have been detected and identified. However, by better constraining the absorption distribution in the faint X-ray selected AGN population, we can hope to answer many questions about the geometry, composition and evolution of the "average" AGN obscuring torus. For example, the strength of the luminosity dependence of the AGN absorption distribution can tell us about how the radiation from the accretion disk influences the surrounding torus, and/or how the torus geometry scales with black hole mass. If the absorption distribution is strongly redshift dependent, then it may be related to the overall evolution of the AGN luminosity function. If the dusty torus provides the reservoir of fuel for the growing black hole, then we might expect to see more obscured AGN in the epochs when AGN were growing most rapidly.

1.4.3 How can the AGN population be better constrained?

The intensity and spectrum of the XRB

Figure 1.9 shows a compilation of several spectral fits to measurements of the XRB made with various telescopes and instruments (Gruber et al. 1999; Gendreau et al. 1995; Lumb et al. 2002, see also Gilli 2003). Despite intensive study, the absolute intensity of the XRB is still rather poorly defined, (only to within $\sim 20\%$). Initial estimates that the deepest *Chandra* observations had resolved over 90% of the soft XRB have been revised because of this uncertainty. However, the spectral shape is rather better determined, especially at lower energies. Between 1 and 20 keV the XRB has a spectral slope of $\Gamma = 1.4$. Below 1 keV, truly diffuse X-ray emission of Galactic origin becomes important, and softens the total spectrum. AGN population models have been devised for the purpose of fitting the spectrum and intensity of

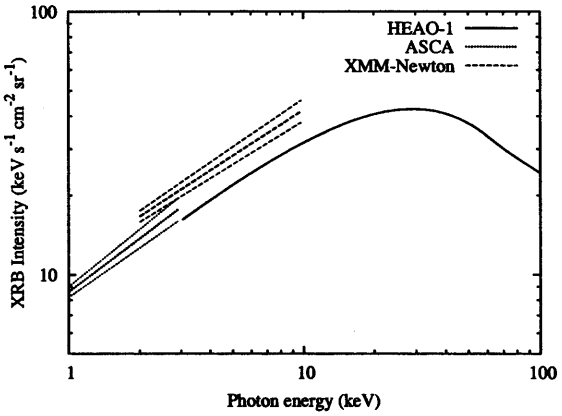

Figure 1.9: The spectrum of the XRB as determined with several X-ray observatories. Fits (with errors where given by the authors) are taken from the literature, and are based on observations with *HEAO-1* (Gruber et al., 1999), *ASCA* (Gendreau et al., 1995), and *XMM-Newton* (Lumb et al., 2002).

the XRB (Comastri et al., 1995; Gilli et al., 2001). However, by itself, the spectrum of the XRB is only a weak constraint.

X-ray source counts

The sky density of sources as a function of flux provides some information about their distribution in redshift and luminosity space.

Two forms of source count measurements are commonly used. The integral source counts, denoted $N(> S)$, are defined as the total sky density of sources with flux above S, whereas the differential source counts, denoted $n(S)$, is the sky density of sources per unit flux interval. The latter is a more powerful measure, because each data point is independent of the others. As a benchmark, one can consider the "Euclidean" case, where the members of some source population are uniformly distributed in Euclidean space, and the luminosity distribution of the sources is unchanging. For such a population we expect to measure $N(> S) \propto S^{-3/2}$ (equivalent to $n(S) \propto S^{-5/2}$). The distance probed for a given flux limit

is proportional to $S_{lim}^{-1/2}$, and the volume surveyed increases with the cube of the distance. If a real source distribution is measured to be steeper than the Euclidean case, then it suggests that the source density increases toward greater distances, and *vice versa.*

The AGN population is obviously far more complex than the simple Euclidean case. AGN were more/less populous at various cosmological epochs, their luminosity distribution has evolved with cosmic time, and they have a range of intrinsic spectral shapes. What is more, the source counts that are observed in a typical faint X-ray sample depend on not just the properties of the underlying AGN population, but also the detection capabilities of the survey. There are a number of biases and selection effects which can affect the shape of the observed source count distribution. The faintest sources in X-ray surveys are typically detected with rather low signal to noise, and only a handful of counts. In this regime, photon counting statistics are important; counting noise makes some sources appear fainter and some brighter. Sources below the nominal flux limit can be detected if they are boosted in this way. If the density of sources just below the flux limit is much greater than the number above the limit, then there can be an apparent surfeit of faint detections. This is the so-called Eddington bias (Eddington, 1913; Teerikorpi, 2004). Source confusion effects can also be important, but only in surveys where the number of detection elements per source is small. These effects are complex, and depend on the shape of the source counts, and the size of the telescope "beam". If the source counts are shallow at the detection limit, then the main effect of confusion will be that faint sources are missed because they are coincident with brighter objects. If the source counts are steep, then the most common confusion effect will be cases where two or more close sources below the detection limit are detected as a single source, leading to an increase in the numbers of faint objects. Each of these effects must be taken into account when calculating the true source counts of any sample.

However, by just considering the source counts in an AGN survey, we can lose a great deal of information about the underlying source population. For example, the

source counts predicted by a population of luminous objects at high redshift can be indistinguishable from the source counts produced by a population of less luminous, but lower redshift objects. For these reasons, examination of the raw AGN source counts provides only a rather weak constraint on the properties of the underlying population. The situation can be improved somewhat by looking at the source count distributions in several energy bands, which, for example, can provide constraints on the relative numbers of absorbed and unabsorbed sources.

X-ray colour distributions

In this study I extensively use X-ray colours as an indicator of absorption in the spectra of faint X-ray selected AGN. Many authors (e.g. Mainieri et al. 2002, Della Ceca et al. 2004, and Perola et al. 2004) have shown that X-ray colour based analyses are effective in deriving the properties of *XMM-Newton* sources which are detected with too few counts to permit full spectral fitting. In the latter studies, the optically identified AGN with and without broad emission lines are seen to occupy separate regions in X-ray colour-colour diagrams. Colour analyses do have limitations. Some information is ignored when considering broad band X-ray colours, because we do not take advantage of the full spectral resolution of, for example, the *XMM-Newton* EPIC detectors. Another problem for AGN samples in which the redshifts are unknown is the degeneracy in X-ray colours between absorbing column density and redshift; a heavily absorbed AGN at high redshift can have X-ray colours indistinguishable from a nearby AGN with only weak obscuration.

Constraints from the direct measurement of the redshift, luminosity and absorption of AGN

By directly measuring the redshift, luminosity and absorption of a X-ray selected sample of AGN to faint flux limits, we use the full information available, and so can put the strongest constraints on the underlying AGN population. However, the size of the samples in which this is possible is almost always small, and is usually

limited by the incompleteness of spectroscopic redshifts for optically faint objects. Photometric redshift methods can help to identify the faint optical counterparts to X-ray sources, but can be biased against objects in the $1.5 < z < 2.5$ range (e.g. Wolf et al., 2004). The optical and X-ray selection functions must be taken into account if we are to estimate the true size of the absorbed AGN population. For example, it is potentially dangerous to assume that the unidentified sources have a similar redshift distribution to the identified objects.

1.4.4 The influence of large scale structure on deep surveys

Surveys of the galaxy population have revealed large density variations on a wide range of length scales (e.g. Zehavi et al., 2002, 2005). These galaxy surveys are typically limited to the relatively low redshift Universe ($z < 1$). However, because of their high luminosities, AGN can be used as a probe of large scale structure at higher redshifts, and clustering of AGN has been measured in optically selected QSO samples, e.g. the 2QZ and SDSS surveys (Croom et al., 2002; Myers et al., 2006). There is also evidence of clustering in X-ray selected AGN samples (Carrera, Fabian, & Barcons, 1997; Carrera et al., 1998; Mainieri et al., 2005). The patches of sky covered by the deepest multi-wavelength surveys are typically $<< 1$ deg^{-2}, raising the possibility that cosmic variance could influence their findings. For example, the deepest single surveys made using *Chandra* and/or *XMM-Newton* sample angular sizes up to $\sim 20 - 30'$. At $z \sim 1$, where we typically find most AGN in deep X-ray surveys, $\sim 20'$ is equivalent to $\sim$10 Mpc. Significant AGN clustering amplitude has been detected at up to several times this scale (e.g. Croom et al., 2002; Myers et al., 2006). It is therefore possible that the redshift distribution measured in these small single fields can be influenced by large scale density fluctuations. In a larger area survey, the spikes and voids in the redshift distribution will be averaged out, and so will provide a better representation of the global AGN population. The moral of this is that one should not rely on a single deep pencil-beam survey to sample the

global average redshift distribution.

1.5 The layout of this thesis

This thesis details the research I have carried out in order to understand the absorption distribution in the AGN population. The work is based primarily on deep observations made by the *XMM-Newton* observatory. In chapter 2, I describe the *XMM-Newton* data reduction process, and source searching methods used throughout this study. I also detail the Monte Carlo simulation techniques I have used to interpret the *XMM-Newton* observations. In chapter 3, I compare the predictions of a number of AGN population models to the distribution of X-ray colours of sources detected in the 13^{H} deep field. I show that some discrimination is possible between the population models using this sample of X-ray sources. In chapter 4, I examine the CDFS field for which deeper *XMM-Newton* data are available, and importantly, the sources have a high completeness of redshift determinations. This removes the degeneracy between redshift and absorption in the colours of X-ray sources. I show that there is no evidence for evolution of the absorption distribution with either redshift or luminosity. In chapter 5, I expand the techniques used in chapter 3, and apply them to a wider sample of X-ray sources detected in a total of six deep *XMM-Newton* fields. There are more than 1800 X-ray sources in this sample. This wider area sample dilutes the effects of cosmic variance, and so defines better the "true" X-ray colour distribution in the AGN population. I compare the predicted X-ray colour distributions of a number of population models to this larger sample. I also examine the amplitude of field-to-field variances in source counts as a measure of cosmic variance. Finally, in chapter 6, I conclude by summarising my findings. I discuss the implications for AGN population models, and the role of obscured AGN growth over cosmic history.

Chapter 2

Data reduction and Monte Carlo simulations

2.1 *XMM-Newton* EPIC data reduction

Here I describe the steps of the data reduction process by which raw *XMM-Newton* data products are eventually transformed into a reliable catalogue of X-ray detections. Broadly speaking, my method for reducing the *XMM-Newton* data for a generic deep field follows the steps described in Loaring et al. (2005). The reductions have been carried out primarily using standard tasks from the *XMM-Newton* Science Analysis Software suite (SAS, version 6.0).

The EPIC data in each of the fields studied in this thesis consist of observations made at several different epochs (in some, spanning more than 18 months). I have not considered AGN variability in this study, and so have combined the observations to produce the deepest possible dataset for each field. In general, the individual observations are taken with different nominal pointings and with a range of position angles. Special considerations for each of the six *XMM-Newton* fields studied in this thesis are detailed in the text later, where most appropriate.

2.1.1 Processing raw *XMM-Newton* EPIC data

Raw data have been obtained from the *XMM-Newton* archive. Calibrated event lists are produced for each observation and for each telescope using the EMCHAIN and EPCHAIN SAS scripts. The EPIC detectors are operated in photon-counting mode, so the important attributes recorded for each event are its time of arrival, raw position, energy channel, and pattern. Energy channels are converted to keV using an appropriate linear relationship. The event pattern records how the energy of a X-ray photon was deposited in the physical pixels of the detector (Ehle et al., 2005). For MOS imaging, only a subset of possible event patterns are considered acceptable (and are calibrated). These are "singles", "doubles", "triples" and "quadruples", i.e. where the photon energy is deposited in one, two, three or four adjacent physical pixels. Other patterns are likely to be non-X-ray induced events, e.g. cosmic ray hits (Kirsch et al., 2005). For the *pn* detector, "single" events only are useful for the softest energy band considered here (0.2–0.5 keV); other patterns introduce high levels of detector noise. For the harder *pn* energy bands ($E > 0.5$ keV), "singles+doubles+triples+quadruples" are used (Kirsch et al., 2005).

2.1.2 Minimising background effects

XMM-Newton suffers from periods of sudden, intense soft proton background flaring, rendering unusable large chunks of observation time. In order to remove background flares from the dataset, the event lists are filtered temporally. A "good time interval" file is created for each observation and camera, which defines the periods where the observed countrate above 10 keV, which traces the particle flux, is not significantly higher than the quiescent levels. The rejection levels have been chosen manually because in general, the quiescent level varies from observation to observation. For some observations, nearly all of the exposure time has to be discarded, For example, in observation 0109661401 (1^{H} field) only $\sim$ 3 ks of good time remain from a total exposure time of 33 ks. Figure 2.1 shows an example of the good time intervals that

were chosen for one of the EPIC observations used in this thesis.

Even in quiescent periods, interactions between cosmic rays and the material of the telescope/detectors cause prominent fluorescence lines at several energies in the EPIC detectors. The impact of these lines is reduced by excluding all events falling in the worst affected detector channels. In MOS the 1.41-1.58, and 1.70–1.82 keV ranges are excluded to remove the Al and Si $K\alpha$ lines. For the *pn*, the ranges 1.36–1.57, 7.38–7.60, and 7.87–9.03 keV ranges are excluded to remove the Al, Cu, Ni and Z $K\alpha$ lines.

The method used to deal with the diffuse background caused by unresolved faint X-ray sources plus residual particle and fluorescence line background is discussed in section 2.1.8.

2.1.3 The EPIC energy bands used in this study

The good spectral resolution and high throughput of EPIC allow the full bandpass to be split usefully into a number of energy bands. The exact choice of these energy bands is a trade-off between sensitivity and the amount of spectral information that can be extracted. For this study, I use four energy bands: nominally 0.2–0.5 keV, 0.5–2 keV, 2–5 keV, and 5–10 keV. The "standard" 0.5–2 keV band is used here for ease of comparison with other studies. Incidentally, the 2 keV upper limit is a good choice because there is a step in the effective area of EPIC at this energy. The 0.2–0.5 keV band covers the remainder of the useful EPIC soft X-ray response. Two bands are used above 2 keV (2–5 keV, and 5–10 keV) rather than a single 2–10 keV band employed by a number of authors. Because the effective area falls with energy, the counts detected from a typical AGN in the 2–10 keV band will be dominated by photons from the softer end of the energy range. The splitting of the 2–10 keV X-ray range has the advantage of allowing an additional measurement of X-ray spectral shape. I note that the XMM-SSC XID team use similar energy bands, but set the division between the two hardest bands at 4.5 keV rather than 5 keV (Watson et

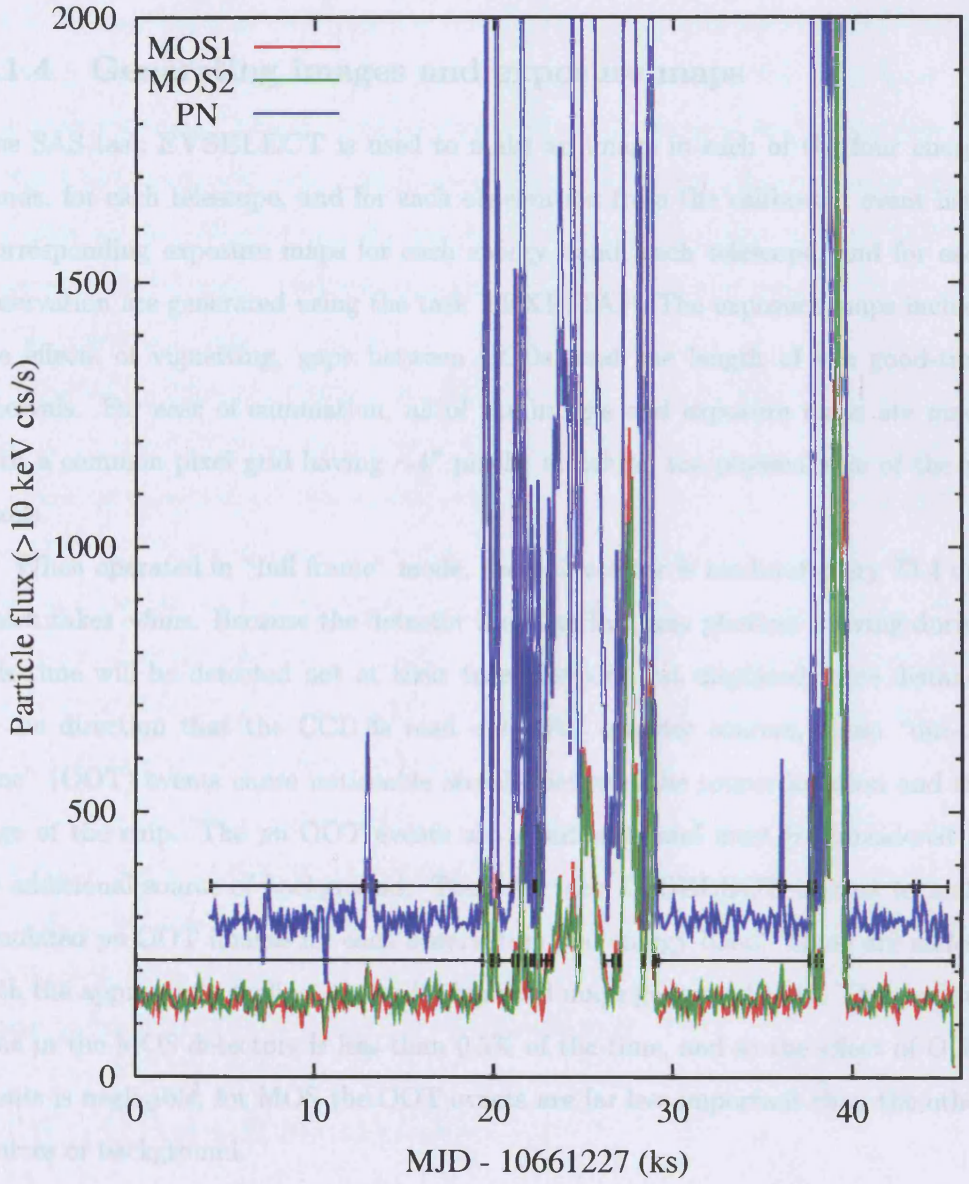

Figure 2.1: Example of the rejection of periods of high particle background during a single 46ks *XMM-Newton* observation of the *Chandra* deep field-North (OBSID 0111550201). The count rate detected above 10 keV is shown, which closely traces the particle background. Horizontal bars indicate the good time intervals.

al., 2001).

2.1.4 Generating images and exposure maps

The SAS task EVSELECT is used to make an image in each of the four energy bands, for each telescope, and for each observation from the calibrated event lists. Corresponding exposure maps for each energy band, each telescope, and for each observation are generated using the task EEXPMAP. The exposure maps include the effects of vignetting, gaps between CCDs, and the length of the good-time intervals. For ease of summation, all of the images and exposure maps are made with a common pixel grid having $\sim 4''$ pixels, matching the physical size of the *pn* pixels.

When operated in "full frame" mode, the *pn* detector is read-out every 73.4 ms, which takes $\sim$5ms. Because the detector is shutterless, any photons arriving during this time will be detected not at their true position, but displaced some distance in the direction that the CCD is read out. For brighter sources, these "out-of-time" (OOT) events cause noticeable streaks between the source location and the edge of the chip. The *pn* OOT events are significant, and must be considered as an additional source of background. The SAS task EVSELECT is used to make simulated *pn* OOT images for each observation and energy band. These are added, with the appropriate scaling, to the background maps generated later. The readout time in the MOS detectors is less than 0.5% of the time, and so the effect of OOT events is negligible; for MOS the OOT events are far less important than the other sources of background.

2.1.5 Astrometric precision of the EPIC detectors

The absolute astrometric precision of EPIC, that is, the conversion from detector coordinates to world coordinates, is rather good (usually better than $\sim 2''$, Ehle et al. 2005), and is stable within each observation. The positions reported by the

three telescopes are mutually consistent. However the absolute astrometric accuracy can be further improved by tying the *XMM-Newton* positions to some well defined reference frame. Four of the *XMM-Newton* fields that I use in this study have also been observed with *Chandra*. The very well defined positions of the sources detected in the *Chandra* observations (which are in turn tied to an optical reference frame) are used as the reference points. For the *XMM-Newton* fields without *Chandra* observations, the optical positions of a number of optically identified QSOs are used as the reference points.

In order to define the required astrometry shifts, the positions of *XMM-Newton* sources in the field must be known. For each observation, the background fitting/source detection process (described fully in section 2.1.8) is carried out on the 0.5–2 keV band images. The offsets between the *XMM-Newton* derived positions and the reference positions are calculated using the SAS task EPOSCORR. These corrections (typically <2″ in RA or Dec) are applied to the world coordinate system (WCS) solution of the original event lists. A new set of images and exposure maps is then made using the revised astrometric solution.

For the purposes of source detection, a set of four (one per energy band) master images are made by summing the individual image frames from MOS1, MOS2 and *pn* and for all the observations of the field. Corresponding master exposure maps are made in each energy band, by computing the *weighted* sum of the individual exposure maps for each detector and each observation.

2.1.6 Additional considerations

A small number of "hot" pixels which are not removed automatically by the SAS pipeline are identified manually. They are then removed by spatially filtering the event lists before remaking the images.

An unexplained enhancement of the 0.2–0.5 keV background level for CCD #5 of MOS1 is evident in a few of the observations used in this study. As a conservative

precaution, for the CDFS field data (see chapter 4), I have excluded the 0.2–0.5 keV MOS1-CCD#5 data for all eight observations. For the LBQS 2212 field the 0.2-0.5 keV MOS1-CCD#5 data are excluded for just the observation in which the enhancement is seen. The images and exposure maps for these observations were remade after these effects were accounted for. A similar background enhancement has been reported by Pradas & Kerp (2005), and is possibly due to scattered light from bright off-axis star(s).

2.1.7 Conversion from EPIC count rate to flux

The energy conversion factor (ECF) is used to convert from a measured source count rate to an absolute flux (in erg s^{-1} cm^{-2}). The ECF is dependent on a number of factors: the detector (*pn* or MOS), the event patterns used, the energy band, the thickness of filter, the Galactic column, and the spectral model assumed for the source. The following method of calculating the ECFs, which follows Loaring et al. (2005), is repeated for the Galactic column densities appropriate for the six *XMM-Newton* fields studied in this thesis.

Firstly, baseline ECFs for the *pn* detector, with the thin filter, have been computed for each energy band. The calculations were carried out using the XSPEC spectral fitting and analysis package (Arnaud, 1996), together with canned on-axis response matrices provided in the standard EPIC calibration. For the 0.2–0.5 keV band, the "singles only" response matrix was used, while for the harder energies, the "singles+doubles" response matrix was used. Fake AGN spectra were generated with XSPEC assuming a power law spectral model, with a photon index of $\Gamma = 1.7$, and Galactic absorption. The predicted countrates were calculated from the fake spectra in each energy band, for exactly the same detector channels used to make the images. The ECF in each energy band is simply the ratio of the predicted countrate to the model flux (before Galactic absorption). The standard EPIC calibration

provides canned response matrices for *pn* "singles" and the "singles+doubles" event-grade sets, but not the "singles+doubles+triples+quadruples" event-grade set used to make the images in this work. However, Osborne (2001) has calculated the increase in count rate expected when "triple" and "quadruple" event grades are also included. The corresponding correction factors are 1.012, 1.055, 1.064, in the 0.5–2, 2-5, and 5-10 keV bands respectively. The final ECFs (including these correction factors) for each of the *XMM-Newton* fields are presented in table 2.1.

As discussed later, the final source detection process is carried out on summed image frames from the MOS1+MOS2+*pn* detectors, and for several separate observations. The response of the MOS detectors is rather different to that of the *pn*. A number of different optical blocking filters have been used which affect the response, especially at soft energies. Therefore, the effective exposure time in each pixel of the summed image is calculated by making a weighted summation of the individual exposure maps frames for each detector and observation. The relative weights of the exposure map frames depend on the relative efficiency of the different detector+filter combinations and are determined as follows. In a similar way to before, XSPEC was used to generate appropriate fake spectra for the MOS detectors with the "thin", "medium" and "thick" filters, and for the *pn* detector with the "medium" filter. Table 2.2 shows the count rates predicted for each of these detector+filter combinations, as a fraction of the *pn*+thin filter countrate. The *pn*+thin filter combination was chosen as the baseline because it has the highest throughput of all the detector+filter combinations.

2.1.8 Background fitting and source detection

An iterative process has been used to characterise the EPIC background, and then to detect sources in the background-subtracted images. This follows the process used in Loaring et al. (2005), and relies on several standard SAS tasks together with a bespoke background fitting program.

Table 2.1: Energy conversion factors (ECF) for the six *XMM-Newton* fields used in this study. The ECF values convert an on-axis *pn*+thin filter count rate into a flux. The assumed spectral model is a power law with $\Gamma = 1.7$ and the Galactic column appropriate for each field. Pattern corrections of 1.012, 1.055, 1.064, in the 0.5–2, 2–5, and 5–10 keV bands are taken from Osborne (2001).

Field	Gal. N_H	ECF (10^{11} cts/s per erg cm^{-2}s^{-1})			
	(cm^{-2})	0.2–0.5 keV	0.5–2 keV	2–5 keV	5–10 keV
13^H	0.6×10^{20}	4.7104	4.9139	1.9611	0.5924
LH	0.6×10^{20}	4.7104	4.9139	1.9611	0.5924
CDFS	0.8×10^{20}	4.6092	4.8878	1.9604	0.5924
CDFN	1.5×10^{20}	4.2758	4.7980	1.9577	0.5923
LBQS 2212	2.4×10^{20}	3.9069	4.6910	1.9544	0.5921
1^H	2.8×10^{20}	3.7489	4.6423	1.9529	0.5921

Table 2.2: The relative response of different EPIC detector/filter combinations. I show ratios of the throughput of the detector/filter combination relative to the throughput of the *pn* detector with the "thin" filter for the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands. The calculations were made assuming a power-law spectrum with slope 1.7, and Galactic absorbing column of 6×10^{19} cm^{-2}.

Tele	Filter	ratio to *pn*+thin			
		0.2–0.5 keV	0.5–2 keV	2–5 keV	5–10 keV
MOS	thin	0.2426	0.3038	0.3702	0.2460
MOS	medium	0.2197	0.2979	0.3661	0.2457
MOS	thick	0.1428	0.2489	0.3515	0.2430
pn	medium	0.8996	0.9791	0.9888	0.9989

The SAS suite provides a pair of tasks (EBOXDETECT and EMLDETECT) for source detection in EPIC images. EBOXDETECT runs a small sliding box over an image to locate regions of enhanced counts, and outputs a list of candidate detections. EMLDETECT takes this shortlist as an input, then carries out a maximum likelihood fit to the position and amplitude of each candidate detection (taking account of the local background and exposure time). Fitting is done in brightness order. Therefore faint sources can be detected even if they lie in the wings of brighter sources. EMLDETECT uses the "small-maps" point spread function model described later. Importantly, EMLDETECT can also be run in multi-band mode, whereby the position and amplitude of each candidate detection is fitted simultaneously in several energy bands. The output of EMLDETECT is a catalogue of positions, countrates/fluxes and detection likelihoods.

However, before the source detection process can be run, a good determination of the EPIC background must be made. As previously discussed, the background in the EPIC images is partly due to particles and this component is fairly uniform across the FOV of the detectors. The remainder of the EPIC background is caused principally by large numbers of astronomical X-ray sources which are fainter than the detection limit of the observations. The latter background component has been focused by the mirrors, therefore it becomes less intense away from the optical axis. An additional source of background in the *pn* images is the OOT events discussed earlier.

The EPIC background model adopted in this study therefore consists of three parts: an unvignetted component, a vignetted component, and the OOT component. The vignetted component is assumed to be some multiple of the exposure map. The unvignetted component is modelled as a multiple of an exposure-mask (the pixels are set to 1 where the exposure time is > 0.01 of the peak, and 0 elsewhere). The OOT component is calculated for the *pn* detector as discussed earlier, but is assumed to be zero for the MOS images.

The following iterative process is used to make a background map in each of the

four energy bands. Firstly, EBOXDETECT and EMLDETECT are used to make an initial sourcelist from the master image, and master exposure map (assuming zero background). This sourcelist is then fed into the background fitting algorithm which is applied separately to the images from each detector and observation. After masking out the locations of sources, the remaining distribution of counts is fitted with a linear combination of the exposure map and the exposure mask. The background fitting program then outputs a map of the best fitting model background. For the *pn* images, the OOT image is included as a fixed component of the fitting process. The individual background maps from each detector and observation are summed to make a master background frame.

The background fitting process is repeated several times; for each iteration the improved background map from the previous iteration is used as an input to the initial EBOXDETECT and EMLDETECT tasks. This background fitting process converges quickly, typically after only two or three iterations. As would be expected, the background levels are approximately proportional to the exposure time of the observations (see figure 2.2). The vignetted background component is broadly similar in MOS1, MOS2 and *pn* once account has been taken for the relative throughput of the detector+filter combinations. However, for a given (*pn*+thin filter equivalent) exposure time, the unvignetted component in the 5–10 keV band is a factor of ~ 2 higher in the MOS detectors than in the *pn*. This is caused by the relatively lower effective area (to X-ray photons) of the MOS detectors above 5 keV; a larger correction is required to convert the MOS exposure times to a pn+thin filter equivalent, than at lower energies.

Once adequate background maps have been fitted, the final source detection run is performed. EBOXDETECT and EMLDETECT are run in multi-band mode on the master images, using the master exposure and background maps for the four energy bands (0.2–0.5, 0.5–2, 2–5, and 5–10 keV). The appropriate ECFs from table 2.1 are passed to EMLDETECT so fluxes may be calculated from the detected count rates. EMLDETECT allows for the adjustment of a number of

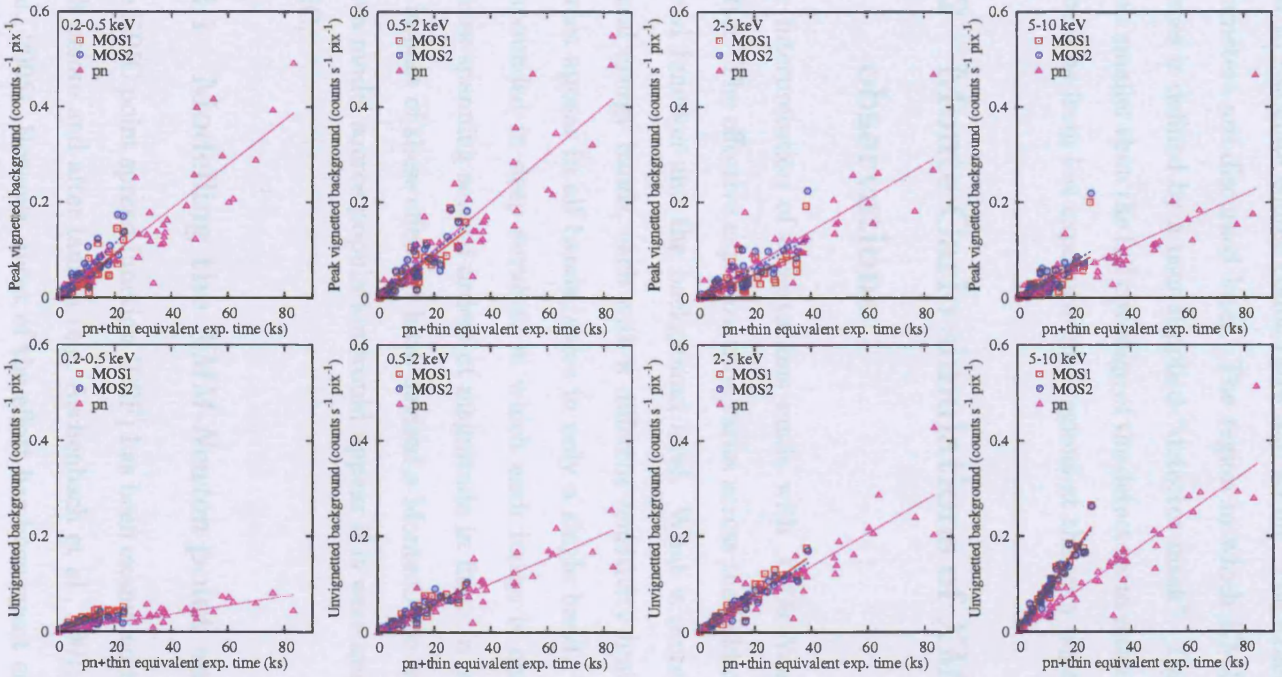

Figure 2.2: Vignetted and unvignetted background components for the observations used in this thesis. The upper row shows the peak background levels as a function of exposure time for the four energy bands. The bottom row shows the same for the unvignetted component of the background. Lines are the best linear fit to the data for each detector. Exposure times have been converted to the *pn*+thin filter equivalent.

fitting parameters. For example, how closely two detections may be placed, the way in which extended sources should be fitted, the maximum radius of the aperture used to calculate source count rates and so on. The exact choices made for these parameters are discussed later. The region in which EMLDETECT searches for sources is defined by a user-supplied "detector mask". The detector mask is a few pixels smaller than the full coverage of the detectors to reduce the number of spurious detections from low exposure-time regions at the very edge of the images.

2.2 Monte Carlo simulations of *XMM-Newton* EPIC observations

The interpretation of observations made with *XMM-Newton* EPIC is a complex matter. The effective exposure time varies across the field of view, as does the point spread function and the background level. What is more, the data are split into several energy bands, each with a different sensitivity limit. Some of the detected sources appear in all bands, some in only a single band. These complications are compounded in deep surveys, in which each image is crowded with hundreds of sources spanning several orders of magnitude in flux. In order to account fully for the impact of these effects, I have devised a Monte Carlo technique which simulates how a model source population would appear if it were imaged with *XMM-Newton* EPIC.

2.2.1 Modelling the *XMM-Newton* point spread function

The EPIC point spread function (PSF) has been examined to various levels of detail, both before and after launch (e.g. Aschenbach et al., 2001; Ghizzardi, 2002; Kirsch et al., 2005). However, most of the effort has been spent characterising the on-axis PSF. Therefore, the current best method to model the PSF at any off-axis angle, and at any energy, is by using a "small-maps" technique. The standard calibration

includes a series of simulated PSF images for a grid of eleven different energies, and seven different off-axis angles. Interpolating these maps gives an appropriate model for the PSF at any off-axis angle and energy. Only one set of small-maps is provided in the EPIC calibration; the differences between the PSFs of the three telescopes are ignored. Apart from the spider-shaped pattern caused by the mirror support struts, the shape of the EPIC PSF is nearly independent of azimuthal angle. Gondoin (2000) found that the small-maps representation of the EPIC PSF was consistent with the in-flight measurements, and thus is adequate for use in a study of faint point-sources such as this. Figure 2.3 shows model PSFs for a range of energies and off-axis angles, compared to actual *pn* images of three real sources (taken from the 1^{H} and 13^{H} deep fields).

2.2.2 Making simulated EPIC images

I have written a program to generate simulated EPIC images. The inputs to this program are an input source list, an exposure map for each detector and energy band, and the EPIC PSF model. The required information for the input sources are their coordinates and count rates in each detector and each energy band.

To start, blank images are created with the same dimensions and pixel scale as the real EPIC images, one for each energy band and detector. The following steps are then carried out for each of the input sources, in each energy band and for each detector. Firstly, the world coordinates of the source are converted to a pixel position in the image. From this, the distance of the source location from the detector optical axis is calculated. A meta PSF-map model for the input source is calculated from the four calibration PSF maps which bound both the "central" energy of the image, and the offaxis angle of the source. The calibration PSF maps are interpolated in energy-space, and then interpolated in off-axis angle space. The shape of the EPIC PSF is only *weakly* dependent on energy (see Fig. 2.3), and therefore not strongly dependent on the exact spectral shape of each model source.

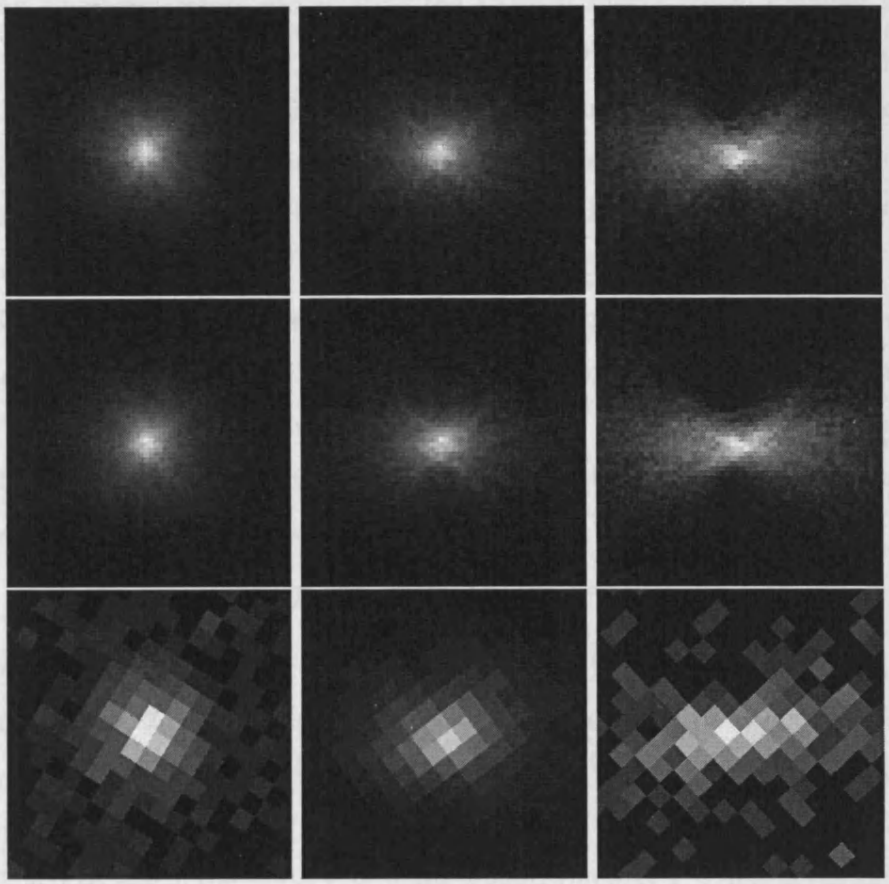

Figure 2.3: Point spread function model for the EPIC detectors, as a function of energy and off-axis angle. The top two rows show the PSF models appropriate for the 0.5–2 keV (top row) and 5–10 keV (middle row) energy bands. Columns show the PSF at off-axis angles of at 3′ (left column), 9′ (centre column), and 15′ (right column). For comparison, the bottom row shows example 0.5–2 keV *pn* images of point-like sources found in the real *XMM-Newton* observations at off-axis angles of 4.0′ (left), 9.3′ (centre), and 14.8′ (right). All images are 66″ on a side. A square-root scale is used to emphasise the shape of the PSF.

So, for simplicity, when interpolating the PSFs in energy space, I simply assume that in each band the X-ray photons all arrive at the log-central energy, i.e. 0.35, 1.0, 3.16, and 7.07 keV, for the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV bands respectively. This is a reasonable approximation considering the relatively narrow energy bands used in this study and the uncertainty on the model PSF provided in the EPIC calibration.

The contribution of the input source to the counts in each pixel of the simulated image is then calculated as the product of the exposure time, the interpolated PSF, and the predicted count rate of the input source.

Therefore, the total counts in each pixel (x, y) of the simulated image for detector $\mathcal{D}$ and energy band $\mathcal{E}$, is given by,

$$C(\mathcal{D},\mathcal{E},x,y) = \sum_i R_i(\mathcal{D},\mathcal{E}) \times \mathcal{T}(\mathcal{D},\mathcal{E},x,y) \times PSF(\mathcal{D},\mathcal{E},x_i,y_i,x,y) \qquad (2.1)$$

where i is summed over all input sources, $R_i(\mathcal{D},\mathcal{E})$ is the predicted countrate of input source i, $\mathcal{T}(\mathcal{D},\mathcal{E},x,y)$ is the value of the exposure map at pixel coordinates (x,y), and $PSF(\mathcal{D},\mathcal{E},x_i,y_i,x,y)$ is the value of the PSF model in pixel (x,y) for a source at pixel coordinates (x_i,y_i).

The pixel size of the model PSF maps is approximately 1/4 that in the simulated images. What is more, the PSF maps are defined at only one azimuthal angle, and so must be rotated. In order to calculate $PSF(\mathcal{D},\mathcal{E},x_i,y_i,x,y)$, the value of the PSF map at the centres of a grid of 3 × 3 sub-pixels is averaged. The translation from the coordinates of the source and image pixels into a location in the PSF map (x_{PSF}, y_{PSF}) is given by,

$$\begin{aligned} x_{PSF}(x_i,y_i,x,y) &= (y_i x - x_i y)/r_i \\ y_{PSF}(x_i,y_i,x,y) &= r_i - (y_i y + x_i x)/r_i \end{aligned} \qquad (2.2)$$

where r_i is the off-axis angle of the source, and the (x_{PSF}, y_{PSF}) = (0,0) at the barycentre of the PSF map.

Once the contribution to the simulated images from all the input sources have been summed, a synthetic background is added to each image pixel. The level and

distribution of this background is discussed later (see section 3.3.1). In order to mimic the effects of photon counting noise, the final number of counts in each pixel of the output images is calculated randomly from a Poissonian distribution about the expectation value.

2.2.3 Modelling the AGN population

In order to make the images described above, a model input AGN population must first be generated. In the process of determining the *XMM-Newton* catalogue for the 13^{H} field, Loaring et al. (2005) used model $N(> S)$ curves to represent the AGN population independently in each of the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands. The simulation process described above was then applied separately in each energy band. While valid for monochromatic studies (e.g. for determining completeness levels or detection likelihood thresholds), the latter technique is not suitable for an X-ray colour analysis (as I use later), because it takes no account of the multi-band properties of individual sources. A better method is to generate a more realistic model population of AGN which have predicted countrates for each energy band. To do this I make the assumption that there exists a single *intrinsic* XLF which describes all AGN, but which is modified by some distribution of absorption to produce the observed XRB, source counts and source colours. During the course of this thesis I use two different model AGN XLFs to represent the intrinsic distribution of AGN in redshift/luminosity space. The details of the population models are described later in chapters 3 and 4.

A simulated population of AGN can be generated by combining the model XLF, a model of the AGN absorption distribution, a model of the average AGN X-ray spectrum, and a model distribution of spectral slopes. To get from this model of the AGN population to a randomly generated sourcelist, the following steps are taken. Firstly the XLF is integrated over a wide range of luminosities and redshifts to calculate the total number of AGN expected in the field. A 1D cumulative

probability distribution is then generated by numerically integrating the 2D XLF via a path which covers the entire z, L_X space. Each point in z, L_X space can then be mapped to a point on the [0,1] interval, and *vice versa.* It is then a simple task to build a list of AGN which are randomly distributed in z, L_X according to the model XLF. Each model AGN is assigned a random value of N_H according to the absorption distribution model, and a random Γ according to the appropriate spectral slope distribution. The simulation process is described in more depth in Chapter 3, including the method I use to convert from AGN parameters to predicted EPIC countrates.

Chapter 3

Constraints on the distribution of absorption using X-ray-selected AGN in the 13^{H} deep field

3.1 Introduction and motivation

Despite the progress made in resolving, and to some extent optically identifying, the hard X-ray population, it has still not been possible to delimit the distribution of absorption in AGN. This problem is particularly acute for the heavily absorbed, high-z AGN, few of which have been detected and optically identified. However, by better constraining the absorption distribution in faint AGN, we can hope to answer many questions about the geometry, composition and evolution of the dusty torus. For example, by measuring if the absorption distribution is luminosity dependent we can hope to constrain how the radiation from the accretion disk influences the surrounding torus, and/or how the torus geometry scales with black hole mass.

This chapter details my investigation of the X-ray properties of sources in a single *XMM-Newton* deep field. I use the X-ray hardness ratio distribution as an indicator of the absorption in the spectra of a sample of $\sim$ 200 X-ray detections, the vast

majority of which are likely to be AGN. The 13^{H} field is in a part of the sky with very low Galactic column density, and has been observed for $\sim$ 200 ks with *XMM-Newton*. As part of an ongoing programme, over a hundred of the counterparts to X-ray sources in the 13^{H} *XMM-Newton* / *Chandra* deep field have been spectroscopically identified. However, most of the optically faint X-ray sources ($R > 23$) remain unidentified, which may mean that some classes of AGN are systematically missed. Therefore, I have taken the X-ray colour distribution of the entire population in the 13^{H} sample, and have compared it to the X-ray colour distributions predicted by a number of AGN population models. It has been shown by numerous authors (e.g. Mainieri et al., 2002; Della Ceca et al., 2004; Perola et al., 2004) that X-ray colour based analyses are effective in deriving the absorption properties of individual *XMM-Newton* sources. This is especially important in the faint source regime where sources are detected with too few counts to permit detailed spectral fitting. However, before undertaking this study it was not clear if the distribution in X-ray colour-colour space of the AGN would retain a measurable "imprint" of the underlying N_H distribution.

In section 3.2 I describe the *XMM-Newton* observations in the 13^{H} field and the source detection process applied to this data set. In section 3.3 I detail the Monte Carlo method I have used to determine optimum detection likelihood thresholds for the catalogue of sources. Then in section 3.3.1 I introduce the simulation scheme used to predict the distribution of sources in X-ray colour-colour space that one would expect to observe with *XMM-Newton* given some model AGN population. The distribution of X-ray colour in the output simulated source populations can be compared directly to the distributions seen in the 13^{H} sample. I examine the predictions of several simple unified scheme models of the AGN population in which the N_H distribution is independent of luminosity and redshift. I also compare the predictions of two more complex model N_H distributions to the 13^{H} sample. In one model, the N_H distribution is dependent on the luminosity of the AGN, and in the other, the N_H distribution is dependent on redshift. In section 3.6.2 I compare the

source count distribution found in the *XMM-Newton* observations of the 13^{H} field to the predictions of the simulated AGN populations. Then in section 3.8 I discuss my findings.

Throughout this chapter I use the following conventions: L_X refers to the *intrinsic* X-ray luminosity of a source in the *observed* 0.5–2 keV band; $S_{0.5-2}$ is the measured flux in the 0.5–2 keV band, corrected for Galactic absorption; N_H is the equivalent hydrogen column density in units of cm^{-2}. I refer to the N_H distribution function as $f(N_H)$, and define it to be the fraction of all AGN, per unit $\log N_H$, which have absorbing column N_H. Most of the material in this chapter has been published in Dwelly et al. (2005). The process I have used to determine the detection likelihood thresholds (see section 3.3) was first presented in Loaring et al. (2005).

3.2 *XMM-Newton* observations in the 13^{H} field

The 13^{H} field was the location of one of the deepest *ROSAT* surveys (McHardy et al., 1998) and was chosen due to its unusually low Galactic absorbing column ($N_H \sim 8 \times 10^{19}$ cm^{-2}), and lack of bright stars. In addition, the 13^{H} field has been the subject of a host of multi-wavelength observations, including a mosaic of four 30ks *Chandra* pointings covering the *XMM-Newton* field of view (McHardy et al., 2003). The field has also been covered by very deep radio mapping at 1.4GHz with the VLA (Seymour, McHardy & Gunn, 2004), and at 610MHz with the GMRT.

The *XMM-Newton* EPIC dataset consists of three observations of the 13^{H} field totalling $\sim$ 200 ks. Multi-band images and exposure maps were created from the raw EPIC data using the method described in section 2.1. After removing periods affected by soft proton flaring, $\sim$ 120 ks of exposure time remained. The sky area covered with at least the equivalent to 50 ks of pn exposure time is ~ 0.15 deg^2. In figure 3.1, I present EPIC images of the 13^{H} field in four energy bands.

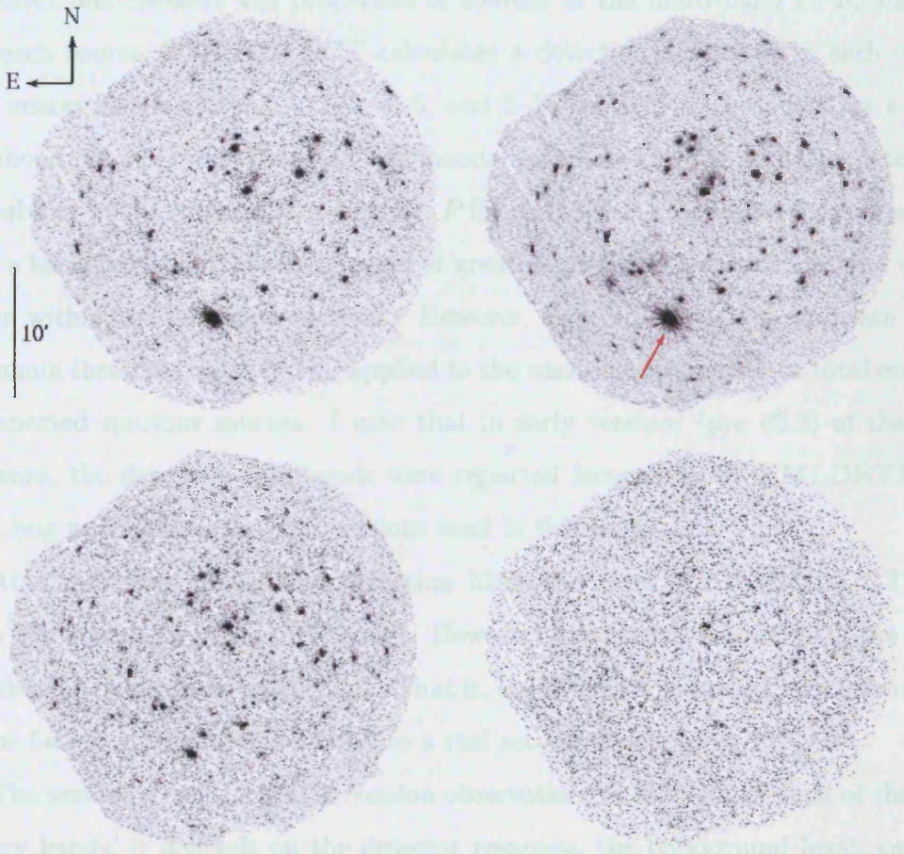

Figure 3.1: The combined EPIC MOS+pn images in four energy bands: 0.2–0.5 keV (top left), 0.5–2 keV (top right), 2–5 keV (bottom left), and 5–10 keV (bottom right). The images are background subtracted, and displayed on a square-root scale. It is evident that the sensitivity is highest in the 0.5–2 keV band. The brightest soft X-ray source in the field is a Galactic star indicated with an arrow.

3.3 Source detection and sourcelist cleaning

The background fitting and source detection process described in section 2.1.8 was applied to the 13^{H} *XMM-Newton* images. The SAS routine EMLDETECT is used to detect and measure the properties of sources in the multi-band EPIC images. For each source, EMLDETECT calculates a detection likelihood in each of the four energy bands (0.2–0.5, 0.5–2, 2–5, and 5–10 keV), $D_{1,2,3,4}$, as well as a total likelihood, D_{tot}. These detection likelihoods are related to the spurious detection probability by the formula $D = -ln(P)$. P for each detected source is the probability that a background fluctuation of equal or greater amplitude to that measured would occur within the detection element. However, it is not trivial to translate some minimum threshold value of D_{tot} applied to the whole sourcelist, into a total number of expected spurious sources. I note that in early versions (pre v5.3) of the SAS software, the detection likelihoods were reported incorrectly by EMLDETECT. This bug was fixed in the SAS versions used in this work.

At a minimum multi-band detection likelihood level of 5.0, EMLDETECT finds 270 detections in the 13^{H} images. However, one expects a fraction of the lower likelihood detections to be spurious. That is, the detection is caused by a fluctuation in the background, and not related to a real astronomical object.

The sensitivity of the *XMM-Newton* observations is different in each of the four energy bands; it depends on the detector response, the background level, and the spectral shape of the sources that one wishes to detect. I have therefore used a Monte Carlo approach to determine the appropriate detection thresholds in each energy band for the 13^{H} *XMM-Newton* dataset. As illustrated by Hasinger et al. (1998), Monte Carlo simulations can be a powerful way to characterise the ability of a survey to detect and measure faint sources near the detection limit. This is especially true in deep X-ray surveys where a number of different processes can contribute to uncertainties in the detected source parameters.

For the purposes of evaluating the detection likelihood thresholds, it is desirable

to simulate a population of sources which is as similar as possible to the real source population. In particular, the simulated sources should have a similar sky density, and a similar flux distribution to the real source population. Effects such as source confusion will then occur in the simulated images at a similar level to the rate at which they appear in the real images.

My simulation method consists of several modular steps that are repeated for a large number of synthetic fields, and carried out separately in each energy band. I briefly summarise the process here, a flow chart of the simulation process is given in figure 3.2. Firstly, an "input" catalogue of sources having random positions and fluxes is generated from a source-count model. Realistic simulated *XMM-Newton* MOS1, MOS2 and *pn* images are generated from this input source population, using the method described in section 2.2, taking account of the exposure time and background levels in the 13^{H} observations. The background fitting and source searching process described in section 2.1.8 is applied to the summed MOS1+MOS2+*pn* images to produce an output sourcelist. A simple pairing algorithm is then used to associate each "output" detection to an "input" source. This process is repeated for a large number of simulated fields to reduce statistical uncertainties in the analysis.

In the following subsections I describe in greater detail the individual steps of my simulation process, and how I have defined the final 13^{H} field *XMM-Newton* sample

3.3.1 Input source populations

Input sourcelists are generated independently for each of the four energy bands. The differential number counts, $N(S)$, are modelled as a broken power law distribution (see equation 3.1). The normalisation constants are chosen such that the differential distribution function is continuous at its knee, i.e. $K_1/K_2 = (S_{knee}/S_{norm})^{\gamma_1-\gamma_2}$.

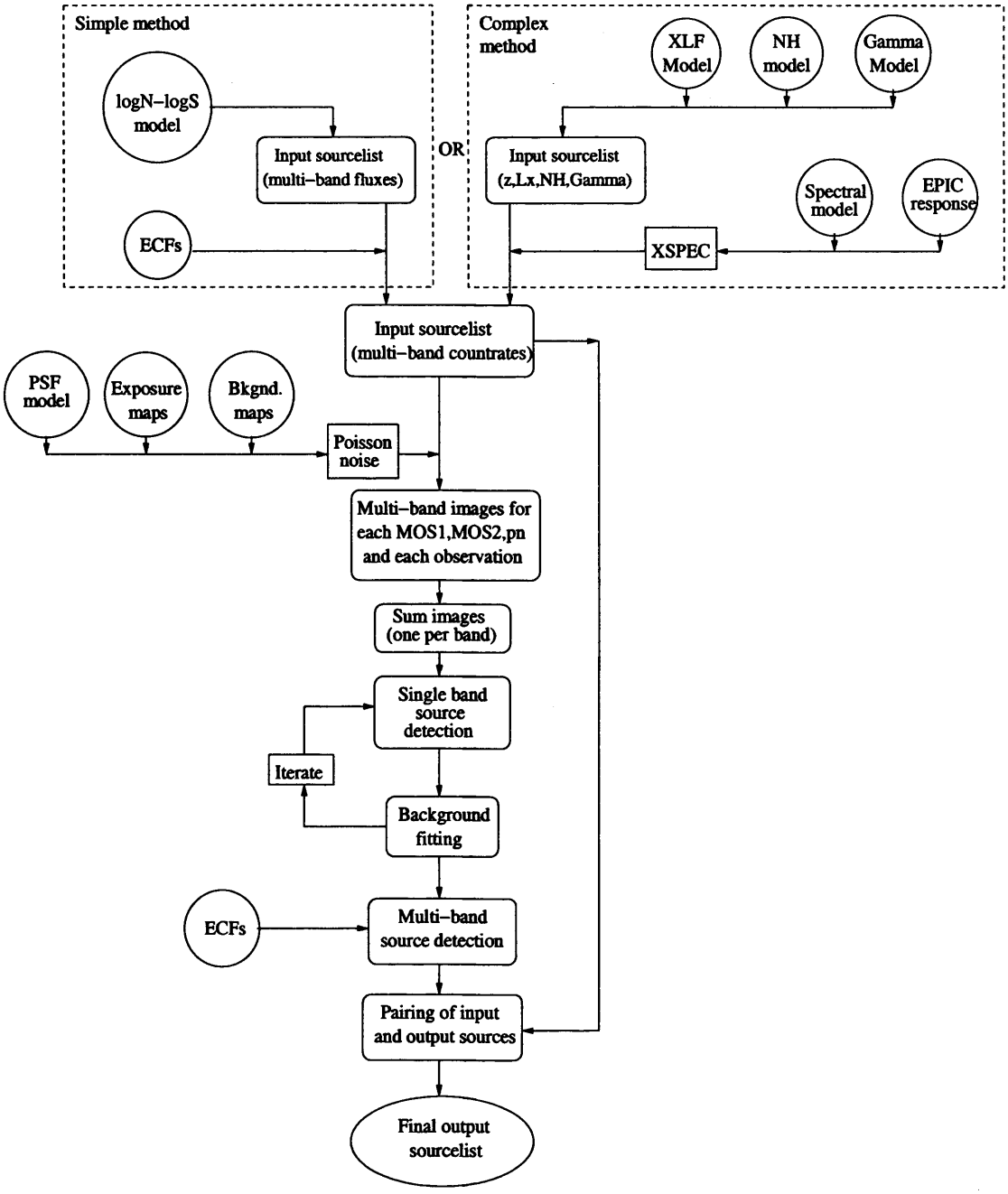

Figure 3.2: A simple flow chart of the AGN population simulation process, broken up into logical steps.

$$n(S) = \begin{cases} K_1 \left(\frac{S}{S_{norm}}\right)^{-\gamma_1} & (S < S_{knee}) \\ \\ K_2 \left(\frac{S}{S_{norm}}\right)^{-\gamma_2} & (S \geq S_{knee}) \end{cases} \quad (3.1)$$

The initial choice of parameters for the source count distribution were taken from fits to the source counts measured in the *Chandra* ultra deep fields (Rosati et al., 2002). However, I found that using this set of parameters, the simulated output source count distribution provided a rather poor match to the source count distribution detected in the 13^{H} sample. Therefore, I iteratively modified the slope and normalisation parameters in order that the simulated source counts provided an acceptable match to the measured distributions.

The input source population is generated for a rectangular area encompassing the entire coverage of the *XMM-Newton* observations. In order to incorporate the effects of source confusion, the lowest flux that is simulated in each band is limited to approximately one fifth of the flux limit reached by the 13^{H} data. The upper limit is constrained to be twice that of the brightest source in the 13^{H} data. The total number of input sources per simulated field, in each energy band, is calculated by integrating the source counts distribution over the allowed flux range. A flux value is randomly assigned to each of these sources from the broken power law $N(S)$ distribution. Each input source is assigned a random position within the EPIC FOV; the effect of source clustering on the detection process is not considered here. Finally, the fluxes are converted to unvignetted count rates via the appropriate ECFs, listed in table 2.1.

3.3.2 Simulated *XMM-Newton* images

The three *XMM-Newton* observations of the 13^{H} field have very similar centres and position angles. By treating the observations as a single observation, it is possible to reduce the amount of processing time required by the simulation process. The exposure maps are simply summed over the three observations for each energy band,

and for each detector MOS1, MOS2 and pn, and treated as if they were a single observation.

Using the input source populations and exposure maps, realistic MOS1, MOS2, and pn images are generated in each energy band following the method described in 2.2. A synthetic background, consisting of a vignetted and an unvignetted component is added to the images. The initial level of the background was set to the level measured in the real 13^{H} images. However, the input source populations reach to fluxes well below the nominal flux limit of the 13^{H} survey, and so the unresolved faint input sources contribute to the total background in the simulated images. To counter this, the level of the added background is iteratively adjusted in order that the *total* background levels match those measured in the real 13^{H} images.

The background fitting/source detection process is carried out on the simulated images following the method of 2.1.8. In order to reduce the amount of processing time, only two iterations of the background fitting process are carried out. An output sourcelist is generated for each energy band by running the EBOXDETECT and EMLDETECT routines in single band mode on the images in each energy band. The output sourcelists are curtailed at a single band detection likelihood of $D_i = 5$.

3.3.3 Matching of input sources and output detections

I have used a simple algorithm to match each output source detected in the simulated images with the appropriate input source. Deciding which input source was actually the "cause" of the output detection is sometimes ambiguous, especially for faint detections. I adopt the input/output source matching method, which accounts for the degradation of the *XMM-Newton* PSF away from the optical axis. I search for input sources within a circular region around the output detection. The radius of the region is $5''$ for sources less than $9'$ off axis, $8''$ for sources between $9'$ and $12'$ offaxis, and $10''$ for sources more than $12'$ offaxis. Where more than one candidate input source lies within the matching region, the input source with the highest flux

is chosen. In figure 3.3 I compare the output fluxes of the simulated sources to the input values.

3.3.4 Determining the detection likelihood thresholds

My goal is to set some "cleaning" criteria to be applied to the real 13^H field "raw" EMLDETECT sourcelist, which removes nearly all erroneous detections, but which leaves the maximum number detections of real astrophysical objects. One must first decide some criteria for a simulated output source to be classed as a valid detection.

Clearly, if there is no input source in the matching region, then the output detection is spurious, and probably caused by a fluctuation in the background. I class all such output sources as "unmatched" detections. The sky density of the input simulated sources is rather high at the faint limit. Therefore, an extremely faint input source could by chance lie within the matching region for an output detection, but not have contributed significantly to the detected flux. For this reason, I also consider to be "unmatched" all those output detections which are matched to an input source having an input flux of less than 0.2 of the output flux.

Some output detections have two or more input sources within the matching region. However, not all of these detections will necessarily be "confused" sources. If for example, there is a large contrast between the fluxes of the input sources, then the output properties (measured by EMLDETECT) will be dominated by the brightest input output. However, in some cases, two input sources of comparable brightness will be detected as a single output source. The output flux will be "boosted" above that of the brightest input source. Following the method of Hasinger et al. (1998), I class output detections to be "confused" if $S_{out}/(S_{inp}+3\sigma_S) > 1.5$, where S_{out} id the output flux, S_{inp} is the flux of the brightest input source, and σ_S is the measurement error on S_{out}. I refer to any detection which meets either the "unmatched" or the "boosted" criteria, to be a "spurious" detection.

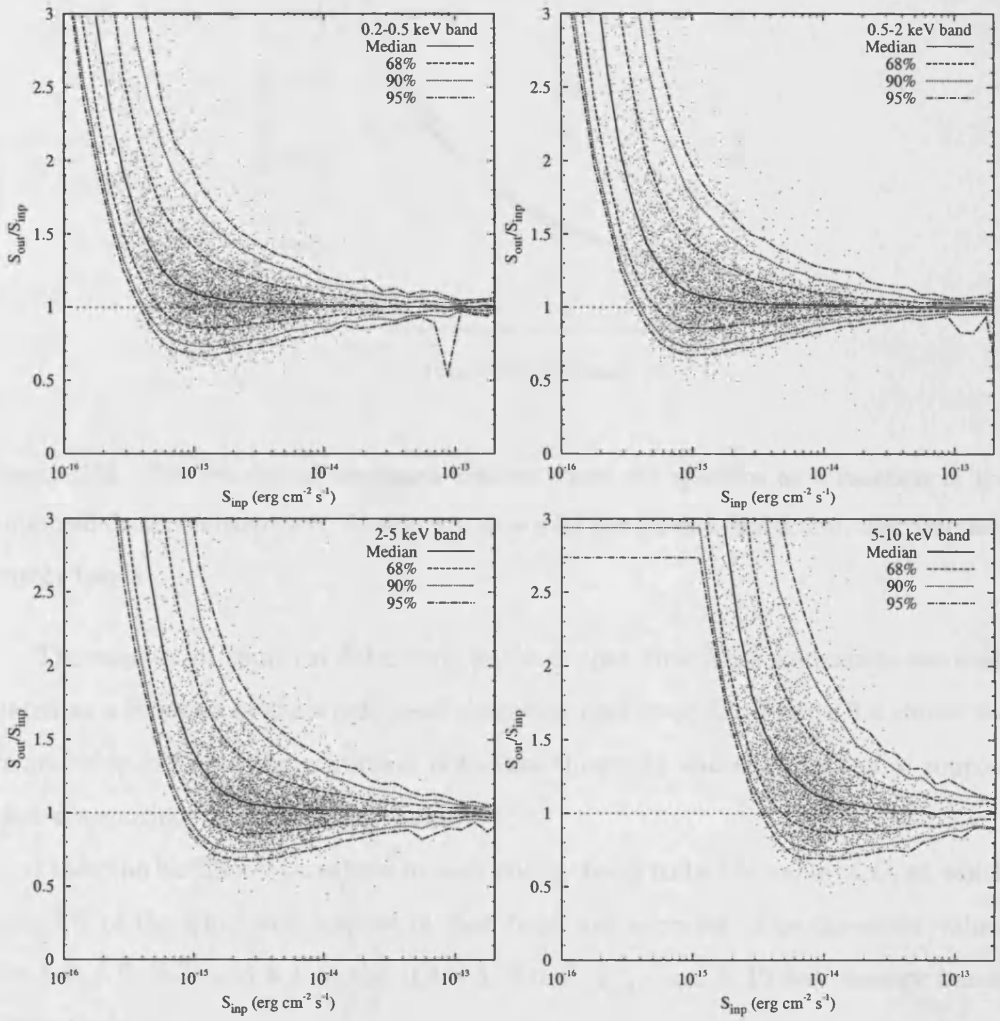

Figure 3.3: The ratio of output flux to input flux for the simulated sources, in each of the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands. The solid line marks the median value as a function of input flux. The dotted, dashed, and dot-dashed lines contain 68%, 90%, and 95% of the sources respectively. Results are shown for simulated populations which have been curtailed at a *single band* detection likelihood of 5.0.

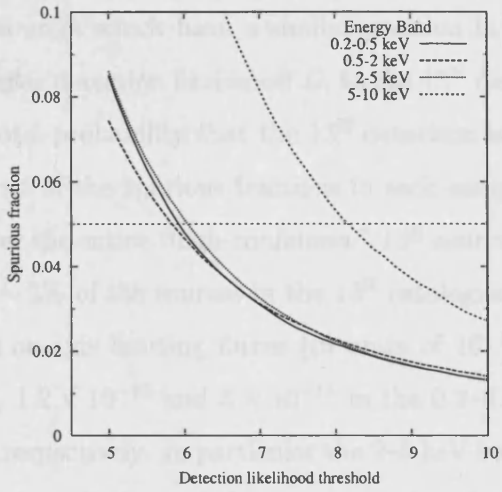

Figure 3.4: The fraction of simulated sources which are spurious as a function of the threshold detection likelihood. Results are shown for the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands.

The number of spurious detections in the output simulated sourcelists are evaluated as a function of the single band detection likelihood D_i. Figure 3.4 shows the relationship between the minimum detection threshold and the fraction of sources that are spurious in each energy band.

I take the likelihood threshold in each energy band to be the value of D_i at which only 5% of the simulated sources in that band are spurious. The threshold values are 5.9, 5.9, 6.0, and 8.1 in the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands respectively.

These thresholds are applied to the sources in the raw 13^H sourcelist. All of the 13^H detections which have detection likelihoods lower than the threshold values in *all* energy bands are removed. Using these criteria 225 sources out of the initial 270 remain in this "high-confidence" 13^H catalogue.

I now estimate the number of spurious detections that remain in the 13^H sourcelist, after these cleaning criteria have been applied. For each of the "high-confidence" 13^H sources, and in each energy band, the spurious fraction is calculated in the

subset of simulated sources which have a similar position in the field (within 2′), and which have a similar detection likelihood D_i to the 13^{H} detection (within 10%). An estimate of the total probability that the 13^{H} detection is a spurious detection is given by the product of the spurious fractions in each energy band. The sum of these probabilities over the entire "high-confidence" 13^{H} sourcelist is $\simeq 7$. Therefore I estimate that only $\sim 3\%$ of the sources in the 13^{H} catalogue are spurious.

The approximate on-axis limiting fluxes (in units of 10^{-15} erg s^{-1} cm^{-2}), are 5×10^{-16}, 5×10^{-16}, 1.2×10^{-15} and 5×10^{-15} in the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands respectively. In particular the 2–5 keV band sensitivity reaches a factor ~ 10 fainter than the knee of the source counts, where the contribution to the XRB, per unit log flux, is greatest.

3.3.5 Removing non-AGN sources from the sample

The purpose of this study is to test the predictions of a number of model AGN populations against the AGN population detected in the 13^{H} field. Therefore it is important for the sample to be as free of non-AGN sources as possible, because these could affect the comparisons. The ongoing optical spectroscopic follow-up program in the 13^{H} field has identified counterparts to over 100 of the *XMM-Newton* sources. In particular, the brightest ($R < 22$) optical counterparts are 92%(81/88) identified. Four of the X-ray sources, including the brightest source in the field, are associated with foreground stars. Therefore I remove these objects from the AGN sample. It is unlikely that many of the remaining X-ray sources with optically faint ($R > 22$) counterparts will be stars.

There are a number of detections that are probably due to emission from clusters in the 13^{H} field. The source fitting task EMLDETECT determined four X-ray sources to have a high likelihood of being extended with FWHM $> 16''$. Of these four sources, three were identified as clusters by Jones et al. (2002) in an analysis of the *ROSAT* imaging of the 13^{H} field. I have removed these four objects from the

13^{H} sample. Jones et al. (2002) determined two additional *ROSAT* detections to be clusters, both of which are also detected in the *XMM-Newton* images. The first of these cluster candidates is determined to be a point source in both the *XMM-Newton* and the *Chandra* imaging of the field, and so is likely to be an AGN. The other candidate cluster is resolved into three separate faint sources in the *XMM-Newton* imaging. It is likely that the relatively poor spatial resolution of *ROSAT* led to the misidentification of these three objects as a cluster. I retain these detections in the final sample.

One might expect a small number of fainter clusters to remain in the sample, but not be flagged as extended by EMLDETECT. To calculate the expected number of such objects, I have made a conservative assumption that the flux limit for detecting extended sources is twice that for point sources. Then by extrapolating the integral source counts plot of Jones et al. (2002) to lower fluxes, whilst incorporating the effective area determination of the survey (Loaring et al., 2005), I find that approximately five clusters will remain in the 13^{H} sample.

After the stars and obvious clusters are removed, the resulting *XMM-Newton* sample contains a total of 217 sources of which the vast majority are likely to be AGN. This is the AGN sample I will use for the remainder of this chapter.

3.4 Modelling the AGN population

To define the source catalogue in section 3.3, I used simple source count models to represent the AGN population independently in each of four energy bands. While valid for monochromatic studies, this technique is not suitable for colour analyses, since it takes no account of the multi-band properties of individual sources. For the purposes of this work I assume that there exists a single intrinsic luminosity/redshift distribution which describes all AGN, and which is modified by some distribution of absorption to produce the observed XRB, AGN source counts and X-ray colours of AGN. I now describe the ingredients of the process I have used to generate input

AGN populations for the simulation process.

3.4.1 Models for the distribution of absorption in AGN

The unified model attributes the X-ray absorption seen in AGN to a dusty torus surrounding the central super massive black hole (Antonucci, 1993). The simplest form of the unified scheme suggests that all AGN are intrinsically similar, and that the observational differences between the various AGN types are primarily due to the orientation of the observer. In such a model, it is the geometry of the dust torus which determines the amount of obscuring material along the observer's line of sight to the central X-ray emitting regions. If it is assumed that all AGN have the same geometry, then it is only the properties of the torus which determine the observed distribution of absorption in the AGN population as a whole. A typical zeroth order approach is to suppose that this characteristic geometry is independent of the luminosity of the central engine, and has not evolved over cosmological timescales. However, alternative scenarios are postulated, (e.g. Gilli, Salvati & Hasinger, 2001; Ueda et al., 2003), which imply more complex forms for the absorption distribution. In this study, I compare the predictions of several different forms of the distribution of N_H in the AGN population ($f(N_H)$).

A very simple form for $f(N_H)$ is a continuous distribution in which the number of AGN per unit $\log N_H$ is proportional to $(\log N_H)^{\beta}$.

$$f(N_H) = \frac{1}{N_{tot}}(\log N_H)^{\beta} \tag{3.2}$$

A similar parameterisation was adopted in the synthesis model of Gandhi & Fabian (2003), who found that setting $\beta = 2$, 5 or 8, gave acceptable fits to the XRB. It should be noted that Gandhi & Fabian (2003) used a separately evolving luminosity function for absorbed AGN and the model used here is not intended to reproduce this. In this model, I constrain the AGN to lie in the range $10^{19} < N_H < 10^{25}$ cm^{-2}, and so $N_{tot} = \int_{19}^{25} f(N_H) d\log N_H = 25^{\beta+1} - 19^{\beta+1}$. I have tested three

models following the $(\log N_H)^{\beta}$ form and refer to them as the $(\log N_H)^2$, $(\log N_H)^5$, and $(\log N_H)^8$ models.

Gilli et al. (2001) introduce two N_H models based on the absorption distribution observed in optically selected Seyfert-2 galaxies (Risaliti, Maiolino & Salvati, 1999), both of which I test in this study. The Gilli et al. (2001) models are parameterised by R, which is the ratio of the number of absorbed to unabsorbed AGN. Model A has R fixed and equal to 4; I refer to it as the *GSH01A* N_H model. Model B of is similar to *GSH01A* at low redshift, with $R = 4$ at $z = 0$, but the ratio of absorbed to unabsorbed AGN increases with redshift up to $R = 10$ for $z \geq 1.32$; I refer to this as the *GSH01B* N_H model. I also test another scenario, similar to *GSH01A*, but with a larger absorbed fraction (i.e. $R = 8$); I call this the *GSH01-R8* model. Note that the N_H distribution of Risaliti, Maiolino & Salvati (1999) contains a number of Seyfert-2s where only the lower or upper limit on absorption is known: Risaliti, Maiolino & Salvati (1999) measured $\sim 4\%$ of Seyfert-2s to have $N_H < 10^{22}$ cm^{-2}, and 26% to have $N_H > 10^{25}$ cm^{-2}. Therefore, in my interpretation of the Gilli et al. (2001) models, $\sim 4\%$ of absorbed AGN lie in the $10^{21} < N_H < 10^{22}$ cm^{-2} interval, and 26% of the absorbed AGN have $N_H = 10^{25}$ cm^{-2}. When applying the R ratio in these models, I consider that all AGN with $N_H > 10^{21}$ cm^{-2} to be 'absorbed'.

I also examine the luminosity dependent N_H model of Ueda et al. (2003), in which high luminosity AGN are more likely to have lower absorbing columns. The form of this N_H function is rather complex, and is stated in terms of the intrinsic rest frame 2–10 keV luminosity, L_{2-10} (in units of erg s^{-1}).

$$f(L_{2-10}, N_H) = \begin{cases} 2 - \frac{5+2\epsilon}{1+\epsilon}\psi(L_{2-10}) & 10^{20} < N_H < 10^{20.5} \text{ cm}^{-2} \\ \frac{1}{1+\epsilon}\psi(L_{2-10}) & 10^{20.5} < N_H < 10^{23} \text{ cm}^{-2} \\ \frac{\epsilon}{1+\epsilon}\psi(L_{2-10}) & 10^{23} < N_H < 10^{24} \text{ cm}^{-2} \end{cases} \quad (3.3)$$

where, the parameter $\psi(L_{2-10})$, which is the fraction of AGN with $10^{22} < N_H < 10^{24}$ cm^{-2}, is given by,

$$\psi(L_{2-10}) = \begin{cases} \frac{1+\epsilon}{3+\epsilon} & \log L_{2-10} < 44 + \frac{1}{\beta}(\psi_{44} - \frac{1+\epsilon}{3+\epsilon}), \\ \psi_{44} - \beta(\log L_{2-10} - 44) & 44 + \frac{1}{\beta}(\psi_{44} - \frac{1+\epsilon}{3+\epsilon}) < \\ & \qquad \log L_{2-10} < 44 + \frac{\psi_{44}}{\beta}, \\ 0 & 44 + \frac{\psi_{44}}{\beta} < \log L_{2-10}. \end{cases} \tag{3.4}$$

The parameter values adopted by Ueda et al. (2003) are $\beta = 0.1$, and $\psi_{44} = 0.5$. The ϵ parameter is the ratio between the number of AGN in the ranges $10^{23} < N_H < 10^{24}$ cm^{-2}, and $10^{20.5} < N_H < 10^{23}$ cm^{-2}. Ueda et al. (2003) fix ϵ to be 1.7 to match the ratio seen in Risaliti, Maiolino & Salvati (1999). The $\frac{1+\epsilon}{3+\epsilon}$ term (= 0.574 for $\epsilon = 1.7$), is the maximum value of the absorbed fraction, $\psi(L_{2-10})$. My simulated populations are defined in terms of the observed frame 0.5–2 keV luminosity, so in order to use the Ueda et al. (2003) model, observed frame 0.5–2 keV luminosities are converted to rest-frame 2–10 keV luminosities using the spectral slope and redshift of each simulated AGN. The following equation is used which incorporates both the difference between the energy bands, as well as the redshift dependent K-correction.

$$L_{2-10} = L_X \times \begin{cases} (1+z)^{\Gamma-2} \times \frac{10^{2-\Gamma} - 2^{2-\Gamma}}{2^{2-\Gamma} - 0.5^{2-\Gamma}} & \Gamma \neq 2 \\ \frac{\ln(10/2)}{\ln(2/0.5)} = 1.161 & \Gamma = 2 \end{cases} \tag{3.5}$$

I refer to the Ueda et al. (2003) N_H distribution as the *U03* N_H model. Note that the *U03* model distribution does not include any AGN with absorbing columns outside the range $10^{20} < N_H < 10^{24}$ cm^{-2}.

Finally, to provide a base-line to the more realistic models, I have investigated scenario in which there are no absorbed AGN; I call this the $R = 0$ model. All of these model N_H distributions are shown in figure 3.5.

To reiterate, for each of the N_H models investigated in this chapter, I take only the $f(N_H)$ part from the published model. I always use the same XLF model to describe the redshift/luminosity distribution of the AGN (see section 3.4.3).

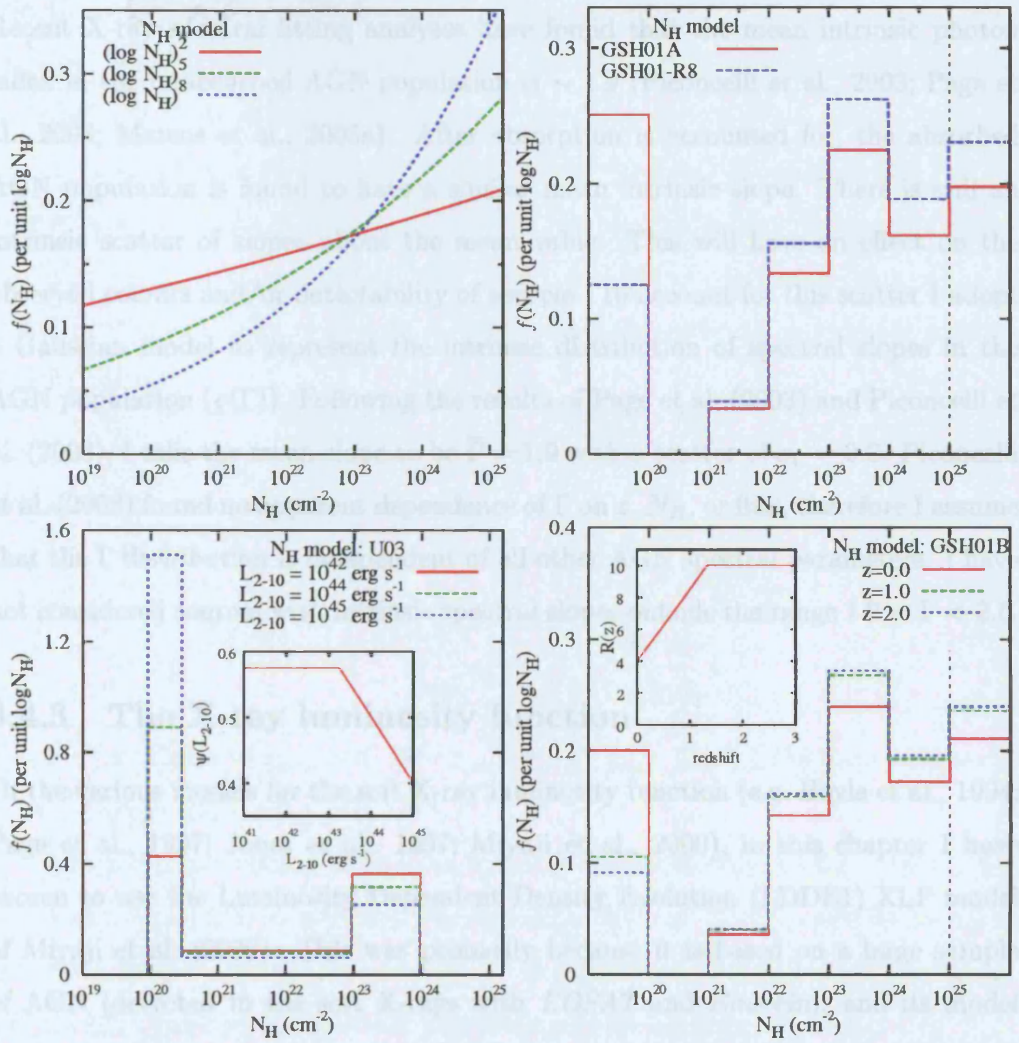

Figure 3.5: The models for the distribution of absorbing column densities which have been examined in this chapter. The plots show the fraction of all AGN which have a particular value of N_H, per unit $\log N_H$. For the *GSH01B* and *U03* N_H models, I also show how the evolution parameter in each model varies with redshift and luminosity respectively. AGN with $N_H < 10^{19}$ cm^{-2} are shown in the leftmost bin, and those with $N_H < 10^{25}$ cm^{-2} in the rightmost bin.

3.4.2 Modelling the distribution of X-ray spectral slopes

Recent X-ray spectral fitting analyses have found that the mean intrinsic photon index in the unabsorbed AGN population is ~ 1.9 (Piconcelli et al., 2003; Page et al., 2003; Mateos et al., 2005a). After absorption is accounted for, the absorbed AGN population is found to have a similar mean intrinsic slope. There is still an intrinsic scatter of slopes about the mean value. This will have an effect on the observed colours and/or detectability of sources. To account for this scatter I adopt a Gaussian model to represent the intrinsic distribution of spectral slopes in the AGN population ($g(\Gamma)$). Following the results of Page et al. (2003) and Piconcelli et al. (2003), I take the mean slope to be $\overline{\Gamma} = 1.9$ with a scatter of $\sigma_{\Gamma} = 0.2$. Piconcelli et al. (2003) found no apparent dependence of Γ on z, N_H, or flux, therefore I assume that the Γ distribution is independent of all other AGN spectral parameters. I have not considered sources with intrinsic spectral slopes outside the range $1.2 < \Gamma < 2.6$.

3.4.3 The X-ray luminosity function

Of the various models for the soft X-ray luminosity function (e.g. Boyle et al., 1994; Page et al., 1997; Jones et al., 1997; Miyaji et al., 2000), in this chapter I have chosen to use the Luminosity Dependent Density Evolution (LDDE1) XLF model of Miyaji et al. (2000). This was primarily because it is based on a large sample of AGN (detected in the soft X-rays with *ROSAT* and *Einstein*), and its model parameters have been determined for the currently preferred lambda-dominated cosmology (Spergel et al., 2003). In section 3.4.3 I discuss the implications of my choice of the Miyaji et al. (2000) XLF model, which has been superceded by other XLF studies since I started the work presented in this chapter.

Miyaji et al. (2000) represent the local soft X-ray luminosity function with a smoothed double power law, defined as,

$$\frac{d\phi(L_X, z=0)}{d\log L_X} = A\left[\left(\frac{L_X}{L^*}\right)^{\gamma_1} + \left(\frac{L_X}{L^*}\right)^{\gamma_2}\right]^{-1} \tag{3.6}$$

where $d\phi(L_X, z = 0)$ is the number density of local AGN with soft X-ray luminosity L_X, per Mpc^3 of comoving volume. L^* is the "knee" of the luminosity function, and defines the transition from the steep slope at high luminosities to the shallower slope at lower luminosities. Miyaji et al. (2000) try several different evolutionary schemes for the luminosity function; pure luminosity evolution (PLE), pure density evolution (PDE), and the more complex luminosity dependent density evolution (LDDE) scheme. The LDDE form was found to provide the best match to the data. The evolution term, $e(z, L_X)$, appears as a multiplier to the zero redshift XLF. That is,

$$\frac{d\phi(L_X, z)}{d\log L_X} = e(z, L_X)\frac{d\phi(L_X, z = 0)}{d\log L_X} \tag{3.7}$$

The evolution term, $e(z, L_X)$, is given by,

$$e(z, L_X) = \begin{cases} 1 & L_X < 10^{-p/\alpha} L_a \\ (1+z')^{p-\alpha\log\frac{L_a}{L_X}} & L_X < L_a \\ (1+z')^p & L_X \geq L_a \end{cases} \tag{3.8}$$

where $z' = z$ for $z < z_{cut}$, and $z' = z_{cut}$ for $z \geq z_{cut}$. The LDDE form reduces to pure density evolution if $\alpha = 0$.

I adopt the best fitting parameter values presented in Miyaji et al. (2000), and where appropriate, have corrected from the $H_0 = 50$ kms^{-1}Mpc^{-1} used by Miyaji et al. (2000), to the currently favoured $H_0 = 70$ kms^{-1}Mpc^{-1}, that is used throughout this study. I give the values of the parameters I have used in table 3.1.

The XLF model, and the $g(\Gamma)$, $f(N_H)$ distribution models define the probability for a simulated AGN to lie in any particular location of z, L_X, Γ, N_H space. The following process is used to generate an input list of sources from the AGN population models. A cumulative probability distribution is calculated by integrating the XLF model via an arbitrary path in z, L_X, and over the range $10^{41} < L_X < 10^{48}$erg s^{-1}, $0.015 < z < 5$. Each point on this cumulative distribution translates to a location in z, L_X space. Therefore, using a random number generator, it is simple to generate randomly distributed positions in z, L_X, where the probability of choosing a

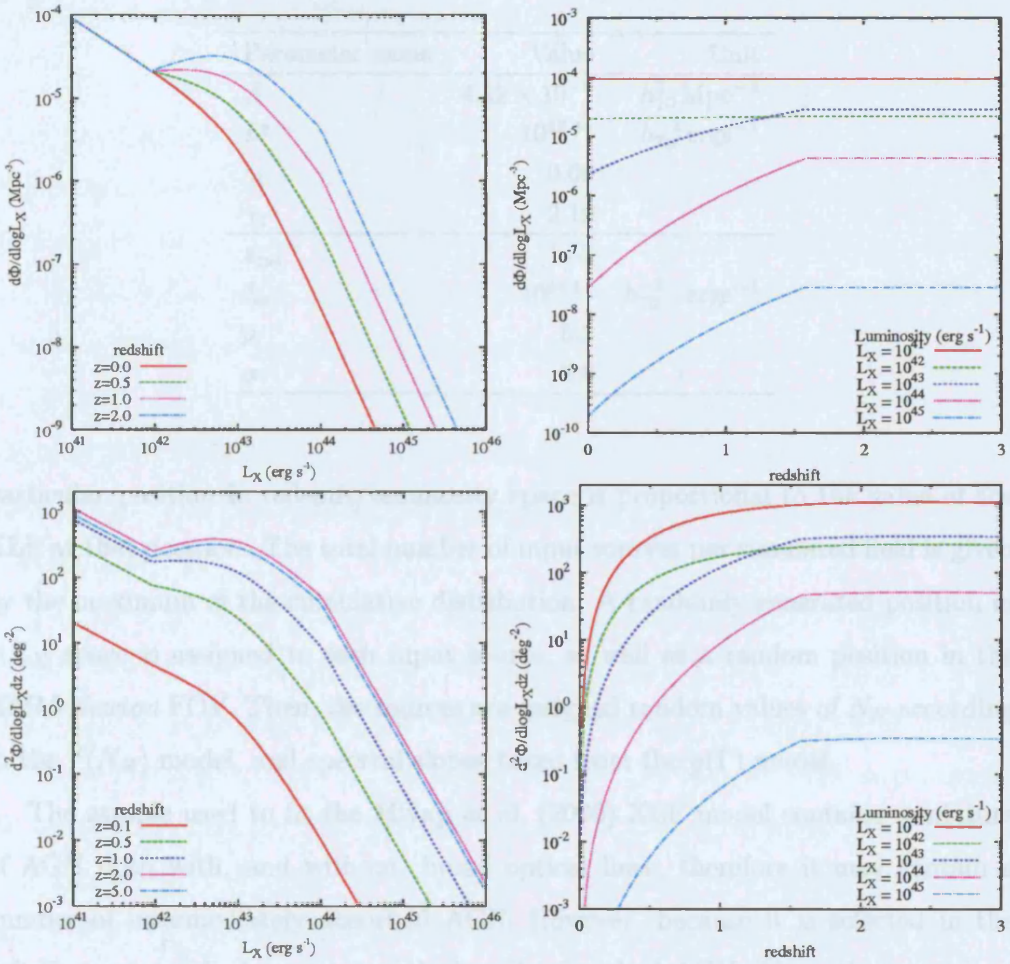

Figure 3.6: The LDDE1 X-ray luminosity function model of Miyaji et al. (2000). The top panels show the luminosity and redshift dependence of the volume density of AGN per unit $\log L_X$. The lower panels show the luminosity and redshift dependence of the sky density of AGN per unit $\log L_X$, per unit redshift, per deg^2. The latter have been calculated by multiplying Φ by the differential comoving volume element (per deg^2) at the appropriate redshift.

Table 3.1: A summary of the LDDE1 X-ray luminosity function parameters of Miyaji et al. (2000), converted to the $h_{70} = 1$ used in this study (where $h_{70} = H_0/70$ km s^{-1} Mpc^{-1}).

Parameter name	Value	Unit
A	4.42×10^{-6}	h_{70}^{3} Mpc^{-3}
L^*	$10^{43.46}$	h_{70}^{-2} ergs^{-1}
γ_1	0.66	
γ_2	2.19	
z_{cut}	1.58	
L_a	$10^{44.11}$	h_{70}^{-2} ,ergs^{-1}
p	5.3	
α	2.6	

particular position in redshift/luminosity space is proportional to the value of the XLF at that position. The total number of input sources per simulated field is given by the maximum of the cumulative distribution. A randomly generated position in z, L_X space is assigned to each input source, as well as a random position in the *XMM-Newton* FOV. Then, the sources are assigned random values of N_H according to the $f(N_H)$ model, and spectral slopes taken from the $g(\Gamma)$ model.

The sample used to fit the Miyaji et al. (2000) XLF model contains a mixture of AGN both with, and without, broad optical lines, therefore it may contain a number of intermediately absorbed AGN. However, because it is selected in the soft-X-rays, sample does not contain heavily absorbed AGN. Thus, the normalisation of the Miyaji et al. (2000) XLF model will underestimate the total (unabsorbed plus absorbed) space density of AGN. I have therefore adjusted the absolute normalisation of the model XLF in order that the simulated output source count distributions match the source counts in the 13^{H} sample at a 0.5–2 keV flux of 2×10^{-15} erg s^{-1} cm^{-2}. The N_H models that I test in this study predict a wide range of ratios of absorbed to unabsorbed AGN, and so a different XLF normalisation is used for each model. For example, for the $(\log N_H)^2$, $(\log N_H)^8$ and *U03* N_H models I respectively used XLF normalisations of 2.4, 3.3 and 5.1 times the normalisation given by

Miyaji et al. (2000).

Consequences for this study of using the LDDE1 XLF of Miyaji et al. (2000)

Since starting the work presented in this chapter, several more sophisticated determinations of the AGN X-ray luminosity function have been published, for example the study of the soft XLF by (e.g. Hasinger, Miyaji & Schmidt, 2005). In addition, XLF models have been calculated from fairly large samples of *hard* (2–10 keV) X-ray selected AGN which, for the first time, reach to reasonably faint fluxes (e.g. Ueda et al., 2003; Barger et al., 2005). These more recent XLF models explore the currently favoured AGN evolutionary picture in which the redshift at which the AGN activity peaks is a function of luminosity, with e.g. the peak of low luminosity AGN activity peaking at $z < 1$.

The Miyaji et al. (2000) XLF model is based on a soft X-ray (0.5–2 keV) selected sample, and so contains primarily unabsorbed "type-1" AGN. The Miyaji et al. (2000) sample is thus unrepresentative of the global AGN population, of which a significant fraction is absorbed. For the purposes of this chapter I have had to make the assumption that the absorbed and unabsorbed AGN have similar distributions in luminosity-redshift space, and have simply "scaled-up" the normalisation of the Miyaji et al. (2000) XLF model to compensate for the missed absorbed AGN.

There are a number of other possible problems with using the Miyaji et al. (2000) XLF model. By selecting sources in the 0.5–2 keV band, the small number of absorbed AGN in the Miyaji et al. (2000) sample might be biased to lie at redshifts where the absorption in their X-ray spectra is redshifted out of the soft X-ray band. The Miyaji et al. (2000) XLF model is defined in terms of "observed" 0.5–2 keV luminosities with no correction made for absorption, thus the luminosities of any absorbed AGN will be underestimated. I have extrapolated the Miyaji et al. (2000) XLF beyond the luminosity-redshift range (specifically to lower luminosities) than are constrained by the Miyaji et al. (2000) sample.

In hindsight it might have been preferable to use one of the more recently published XLF models rather than the Miyaji et al. (2000) LDDE1 model to represent the redshift-luminosity distribution of the AGN populations modelled in this chapter. However, the main aim of this work is to compare the predictions of a number of AGN absorption distribution models using their predictions for the X-ray colour distribution, which, as shown later in chapter 5, is not strongly dependent on the underlying XLF model. Therefore the decision to adopt the Miyaji et al. (2000) LDDE1 XLF model should be seen as only a minor caveat for the results of this chapter.

3.4.4 X-ray colours from AGN spectral templates

At this stage in the simulation process, the input sources have each been assigned values of z, L_X, Γ and N_H. In order to create simulated images from these source parameters, I have to convert them into expected EPIC countrates. I have determined the X-ray colours of the simulated AGN by using a simple absorbed power-law (APL) model, which also includes a correction for the small Galactic absorbing column ($N_H \sim 8 \times 10^{19}\mathrm{cm}^{-2}$) found in the 13^{H} field. I use the spectral fitting package XSPEC (Arnaud, 1996) to generate fake spectra, incorporating both the instrumental response (for the MOS and *pn* cameras), and the AGN parameters z, N_H and Γ. The fake spectra are then summed over the appropriate energy bands, using the same ranges of detector channels that are used to make the EPIC images. This allows a prediction of the countrate that one would expect to see in EPIC for an AGN with some particular set of spectral parameters. However, the cost in processing time would be prohibitive if the simulation process individually recalculated these conversion factors for each of the thousands of simulated AGN. To get around this, I have built lookup tables of conversion ratios, which finely sample (z, N_H, Γ) parameter space, covering the range $0.01 < z < 5$, $10^{19} < N_H < 10^{25}$ cm^{-2}, $1.2 < \Gamma < 2.6$.

I have also examined the effect of including a small component of reflected emission (reflected off cold material) in the spectral model (see 1.2.2). The physical interpretation is that a fraction of the X-rays from the central engine are directed into our line of sight after being reflected off the inner face of the absorbing torus. For unabsorbed and intermediately absorbed AGN, the addition of the reflection component has the net effect of hardening the spectrum at high energies. For AGN in which the primary power law is very heavily absorbed, the soft X-rays from the reflected component can be the only soft X-ray emission we see from the AGN.

In order to test the predictions of this model, I have used XSPEC to make a set of lookup tables as before, but this time with the *pexrav* model of Magdziarz & Zdziarski (1995). I have used the following *pexrav* parameters: the reflecting material covers π steradians of the sky, the system is viewed at an inclination angle of 30 deg, and the reflecting material has solar elemental abundances. I refer to this as the "absorbed powerlaw plus reflection" (APL+R) spectral model.

It is beyond the scope of this study to include more complex AGN spectral features, such as FeK lines, or scattered soft X-ray emission. I expect the effect of these features on AGN colours to be small relative to the effects of continuum obscuration. However, as shown later, the results suggest that a number of the absorbed sources in the 13^{H} field do have a significant soft X-ray component. Although FeK lines are seen in many high signal-to-noise X-ray spectra of AGN, the measured equivalent widths are almost always less than a few hundred eV.

It is not expected that the 13^{H} observations will detect very many (if any) very heavily absorbed AGN ($N_H > 10^{25}$ cm^{-2}), and so I have not included such objects in the simulated populations. In fact, the simulations predict that AGN with $N_H > 10^{24}$ cm^{-2} should account for only $\sim 1\%$ of the detections in the 13^{H} sample. Therefore, any additional attenuation due to Compton scattering within the dusty torus is ignored, since it will have little effect for the bulk of the 13^{H} sources (which have $N_H < 10^{24}$ cm^{-2}).

The lookup tables are used to convert rapidly from any set of simulated AGN

parameters L_X, z, N_H, Γ to count rates, in each energy band, and for each EPIC detector. The exact count rates for each simulated input source are calculated from the appropriate lookup table using a bi-linear interpolation of the conversion values which bound the desired location in z, N_H space.

3.4.5 Carrying out the Monte Carlo simulations

The process described in section 3.3.2 is used to generate simulated EPIC images from the model input populations, and then to carry out the background fitting/source searching process on the simulated images.

I have carried out a set of simulations separately for each of the $f(N_H)$ models described in section 3.4.1. For each $f(N_H)$ model, 100 fields worth of images were simulated and then source searched. The simulation process was carried out firstly using the absorbed powerlaw spectral model, and then again using the "absorbed powerlaw plus reflection" spectral model. The simulation process consumes considerable processing time, particularly the SAS source searching tasks. The resultant simulated output sourcelists were curtailed with the same detection likelihood criteria as were applied to the real 13^{H} sourcelist (see section 3.3).

I decided to run 100 fields worth of simulations for each combination of N_H model and spectral model. This number of simulations provided an adequately large population of simulated sources for comparison with the real 13^{H} sources, without consuming an unacceptably large amount of processing time.

3.5 Capabilities of the 13^{H} survey

The inherent capabilities and limitations of the *XMM-Newton* data in the 13^{H} survey field can be precisely evaluated from the simulated source populations. These simulations can also be used to predict the distribution of X-ray properties of the sources in the 13^{H} sample that would be expected given a number of AGN population models. What is more, one can evaluate how sensitive a comparison in X-ray

colour space is to different AGN absorption distribution models.

I have used the method described in section 3.3 to match the simulated output detections to input sources. Then, for example, the X-ray colours of the output sources can be related to the input spectral parameters (z, L_X, N_H, and Γ).

3.5.1 X-ray selection function

To determine the selection function of the survey, the fraction of output detections with input sources is evaluated as a function of the input parameters z, L_X, N_H, and Γ. Figure 3.7 shows the 50% completeness limit, as a function of redshift and luminosity, for several different levels of absorption. The contours show the loci in $z - L_X$ parameter space at which half of the input sources have output detections. There is a clear reduction in detection probability for AGN having absorbing columns above $10^{23.5}$ cm^{-2}, and this effect is more marked at low redshift. The plot shows that the 13^{H} survey is able to detect more than 50% of any AGN which are luminous ($L_X \geq 10^{44}$ erg s^{-1}) and have moderate-absorption ($10^{21.5} < N_H < 10^{22.5}$ cm^{-2}) up to $z \approx 3.5$.

Figure 3.8 shows the fraction of all input sources that are matched to output sources in the simulated images, as a function of their degree of absorption. This illustrates the differences in detection probability between the two spectral models; the addition of a reflection component to the spectral model has a small effect on the detectability of heavily absorbed sources. The dependence of the selection function on Γ is also shown in figure 3.8, which shows the fraction of simulated input sources with output counterparts, as a function of their spectral slope. It can be seen that the spectral slope of an AGN has a small but measurable bearing on its probability of detection. A strong increase in detection probability is seen for very hard slope sources ($\Gamma < 1.4$), however, the inset histogram shows that very few of these objects are predicted by the $g(\Gamma)$ model. I have used the 0.5–2 keV de-absorbed flux to normalise the model spectra, so the hard-slope AGN have a

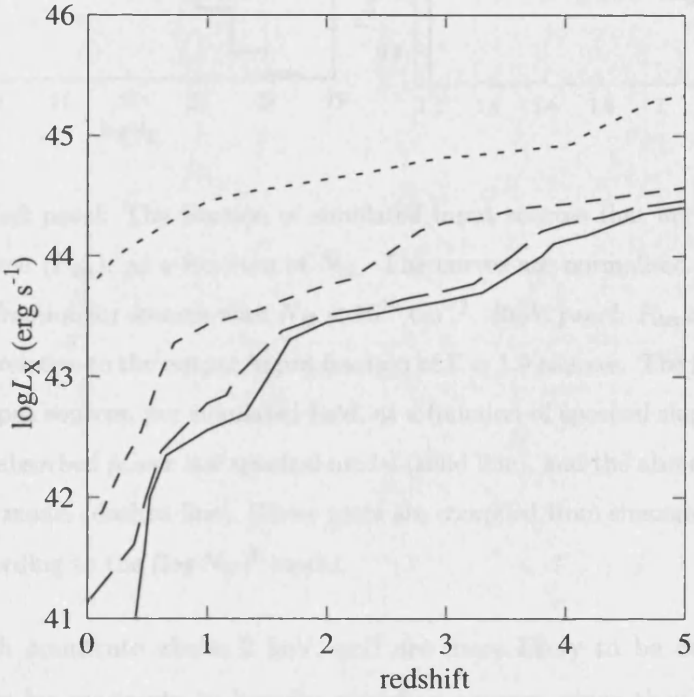

Figure 3.7: The 50% detection limits of the 13^{H} *XMM-Newton* dataset. The 50% detection limits have been calculated from the simulated source populations, as a function of redshift and intrinsic observed frame 0.5–2 keV luminosity. The results for four ranges of absorption are shown: $0 < N_H < 10^{21.5}$ cm^{-2} (solid), $10^{21.5} < N_H < 10^{22.5}$ cm^{-2} (long dash), $10^{22.5} < N_H < 10^{23.5}$ cm^{-2} (short dash), and $10^{23.5} < N_H < 10^{25}$ cm^{-2} (dotted). The results are for the simulated AGN population which was generated using the absorbed power-law spectral model and with the $(\log N_H)^8$ model.

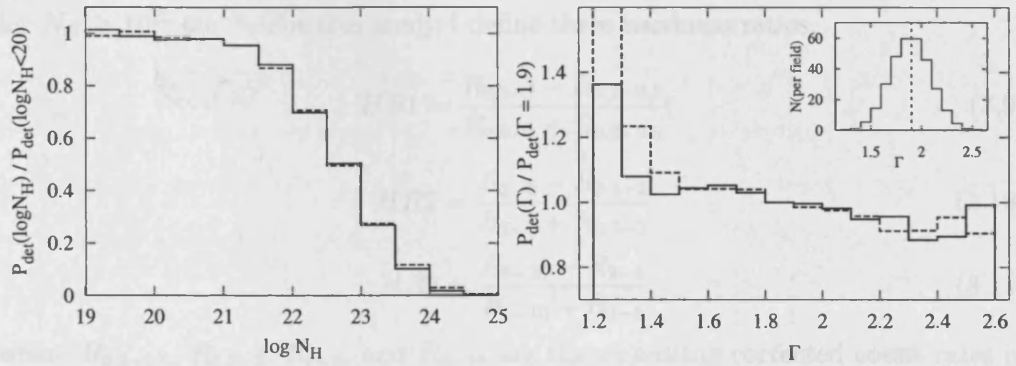

Figure 3.8: *Left panel:* The fraction of simulated input sources that are matched with an output source (P_{det}), as a function of N_H. The curves are normalised relative to the output/input fraction for sources with $N_H < 10^{20}$ cm^{-2}. *Right panel:* P_{det} as a function of Γ, normalised relative to the output/input fraction of $\Gamma = 1.9$ sources. The inset shows the number of output sources, per simulated field, as a function of spectral slope. Results are shown for the absorbed power law spectral model (solid line), and the absorbed power law plus reflection model (dashed line). These plots are compiled from simulated populations generated according to the $(\log N_H)^8$ model.

relatively high countrate above 2 keV, and are more likely to be detected. This effect is larger for moderate to heavily absorbed sources, since they are primarily detected at these harder energies. The impact on the overall selection function is largest for the $f(N_H)$ models containing the largest fraction of absorbed sources, i.e. the *GSH01-R8* and *GSH01B* N_H models.

3.5.2 Sensitivity of X-ray colours to absorption

Constraints on $f(N_H)$ models can be made from analysis of X-ray colour distributions. For example, Perola et al. (2004) compared the N_H of *XMM-Newton* sources determined from full spectral fits, with the N_H estimated using a hardness ratio method (over the 0.5–10 keV range), and showed that they were in good agreement

for $N_H \geq 10^{22}$ cm^{-2}. For this study I define three hardness ratios,

$$HR1 = \frac{R_{0.5-2} - R_{0.2-0.5}}{R_{0.5-2} + R_{0.2-0.5}} \tag{3.9}$$

$$HR2 = \frac{R_{2-5} - R_{0.5-2}}{R_{2-5} + R_{0.5-2}} \tag{3.10}$$

$$HR3 = \frac{R_{5-10} - R_{2-5}}{R_{5-10} + R_{2-5}} \tag{3.11}$$

where $R_{0.2-0.5}$, $R_{0.5-2}$, R_{2-5}, and R_{5-10} are the vignetting corrected count rates in the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands respectively. The corresponding measurement errors are denoted by σ_{HR1}, σ_{HR2}, and σ_{HR3} and are calculated from the errors on the countrates thus,

$$\sigma_{HR1} = 2\frac{\sqrt{(\sigma_{R0.5-2}\,R_{0.2-0.5})^2 + (\sigma_{R0.2-0.5}\,R_{0.5-2})^2}}{(R_{0.5-2} + R_{0.2-0.5})^2} \tag{3.12}$$

$$\sigma_{HR2} = 2\frac{\sqrt{(\sigma_{R2-5}\,R_{0.5-2})^2 + (\sigma_{R0.5-2}\,R_{2-5})^2}}{(R_{2-5} + R_{0.5-2})^2} \tag{3.13}$$

$$\sigma_{HR3} = 2\frac{\sqrt{(\sigma_{R5-10}\,R_{2-5})^2 + (\sigma_{R2-5}\,R_{5-10})^2}}{(R_{5-10} + R_{2-5})^2} \tag{3.14}$$

The count rates, hardness ratios, and their respective errors are computed within EMLDETECT using the combined dataset from the MOS1, MOS2 and pn detectors. If any hardness ratio measurement is undetermined (due to zero countrates in two consecutive energy bands), I set it to 0.0 ± 1.0.

The dependence of $HR1$, $HR2$ and $HR3$ on absorption is illustrated in figure 3.9, which shows the measured X-ray colour-colour distributions of output simulated sources in four bins in N_H. For each of these N_H bins, I have over-plotted the expected track in colour-colour space for a model absorbed power law spectrum with absorption in the log-centre of the bin, $\Gamma = 1.9$, and a range of redshifts. The width of the N_H bins, and the range of Γ in the simulated sources act to distribute sources about the model track. The relative density of the distribution along the track is mostly determined by the evolution of the XLF model, which in the case of the LDDE1 XLF, peaks at $z \sim 1.5$. A significant amount of scatter is caused

by measurement uncertainties within the source detection process, particularly for the faintest sources. For the purposes of this illustration, I have removed from the simulated catalogues the small number of output sources which have no valid input candidates within the matching region. However, later, when I compare the simulated AGN populations to the 13^{H} sample, these output detections are retained. This is because I expect a small number of spurious detections to remain in the 13^{H} sample.

One can see from the left-most upper panel of figure 3.9, that the colour distribution of output simulated sources with $N_H < 10^{21.5}$ cm^{-2} is compact, and approximately centred on $(HR1, HR2) = (0.2, -0.5)$. The study by Mainieri et al. (2002), of *XMM-Newton* sources detected in the Lockman hole, found that of the AGN lying in this part of hardness ratio space, the vast majority had optical counterparts with broad lines, and had X-ray spectra with little or no X-ray absorption ($N_H < 10^{21.5}$ cm^{-2}). In contrast, most of the identified AGN having hard X-ray colours in the Mainieri et al. (2002) sample had narrow line optical counterparts, although only a small fraction of their hard sources had optical identifications. Examination of the three upper right panels of figure 3.9 reveals that the moderately to heavily absorbed sources ($N_H > 10^{21.5}$ cm^{-2}), occupy a measurably different region of $HR1$,$HR2$ space compared to their less absorbed counterparts. In particular, for the simulated AGN population shown here, $HR1$ is sensitive to absorption in the range $10^{21.5} < N_H < 10^{23.5}$ cm^{-2}, and $HR2$ is sensitive to absorption above $N_H = 10^{22.5}$ cm^{-2}. In the study of Georgantopoulos et al. (2004), the hardness ratio between the 0.5–2 and 2–8 keV bands did not appear to separate the broad and narrow line AGN; however, relatively few of the harder AGN in this sample had spectroscopic identifications. Della Ceca et al. (2004) showed that the majority of AGN with broad line counterparts fall in the range $-0.75 < HR2 < -0.35$, consistent with the location of the low absorption AGN ($N_H < 10^{21.5}$ cm^{-2}) seen in these simulated populations. Caution should be taken when making direct comparisons of the hardness ratios of AGN found in these studies. For example, the HR values

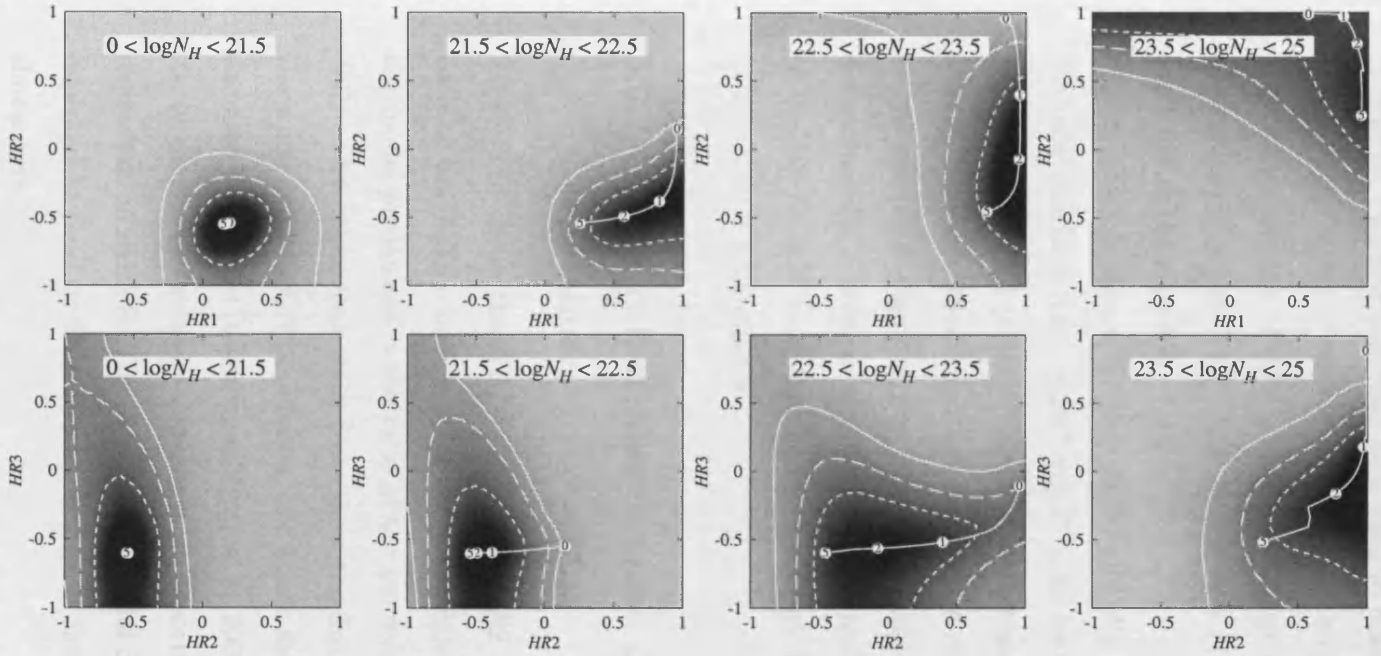

Figure 3.9: X-ray colour-colour plots showing the sensitivity of hardness ratios to absorption. The plots show the $HR1$ *vs* $HR2$ (upper row) and $HR2$ *vs* $HR3$ (lower row) distributions for simulated sources modelled with absorbed power law spectra. Each panel shows the HR distribution of output simulated sources grouped according to their intrinsic absorption ($0 < N_H < 10^{21.5}$ cm^{-2}, $10^{21.5} < N_H < 10^{22.5}$ cm^{-2}, $10^{22.5} < N_H < 10^{23.5}$ cm^{-2}, and $10^{23.5} < N_H < 10^{25.0}$ cm^{-2}). The levels of the contours are set such that they include 50% (short dash), 75% (long dash), and 90% (solid line) of the sources. The contribution of each simulated source to the greyscale map was represented by an elliptical Gaussian centred on the measured position in colour-colour space, and having widths equal to the corresponding σ_{HR}. The white tracks show the locus of expected X-ray colours for an AGN with an absorbed powerlaw spectrum, N_H in the logarithmic centre of the interval, $\Gamma = 1.9$, and with redshift ranging from 0 to 5 (numbered points indicate z).

depend on the Galactic column, the EPIC filters that were used, and the exact set of detector channels that have been used to make the images.

As might be expected, many of the faint output simulated sources do not have good measurements of $HR3$. Such output sources have noise dominated countrate measurements above 2 keV, and hence have $HR3$ measurements fairly randomly scattered in the interval $[-1, 1]$.

As shown in figure 3.7, for AGN with $N_H > 10^{23.5}$ cm^{-2}, one expects only the most luminous ($L_X > 10^{44}$ erg s^{-1}), to be detectable in the 13^{H} survey. The bottom right hand panel of figure 3.9 illustrates that $HR3$ is sensitive to these high column densities for all but the highest redshift AGN. Indeed, in the bright source sample of Caccianiga et al. (2004), hard band X-ray count rates were well determined for a number of objects; of the four objects with a higher count rate in the 4.5–7.5 keV band than in the 2–4.5 keV band, all had narrow emission line galaxy optical counterparts.

3.5.3 X-ray colour distributions produced by model AGN populations

Figure 3.10 shows the $HR1$ *vs* $HR2$ and $HR2$ *vs* $HR3$ distributions predicted by each of the $f(N_H)$ models. The most immediately noticeable difference between the model predictions is the ratio of the number of sources which lie either side of $HR1 = 0.6$. Note that these plots show the distribution in hardness ratio space of the whole model AGN population, and hence are dependent to some degree on the assumed XLF model (taken from Miyaji et al. 2000). However, as shown in figure 3.9, X-ray colours are strongly dependent on absorbing column, and relatively weakly dependent on redshift. Therefore, a substantial change in the model XLF would be expected to produce only a small change in the model colour-colour distributions shown here.

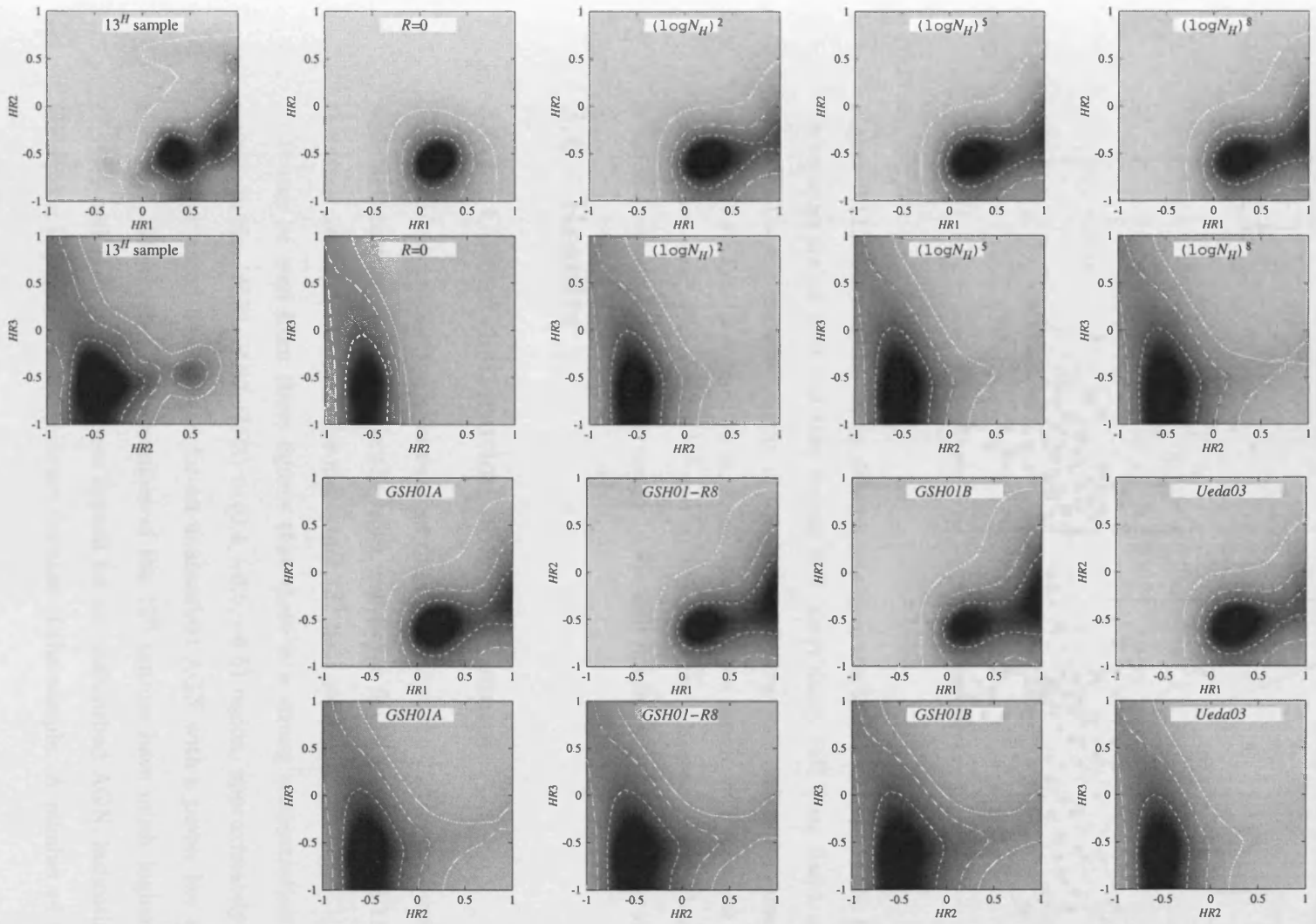

Figure 3.10: The X-ray colour-colour distributions predicted by the $f(N_H)$ models compared to the distributions observed in the 13^H sample. The results are for model populations generated according to the absorbed power-law spectral model. The levels of the contours are set such that they include 50% (short dash), 75% (long dash), and 90% (solid line) of the smoothed source distribution, and were generated in the same way as for figure 3.9.

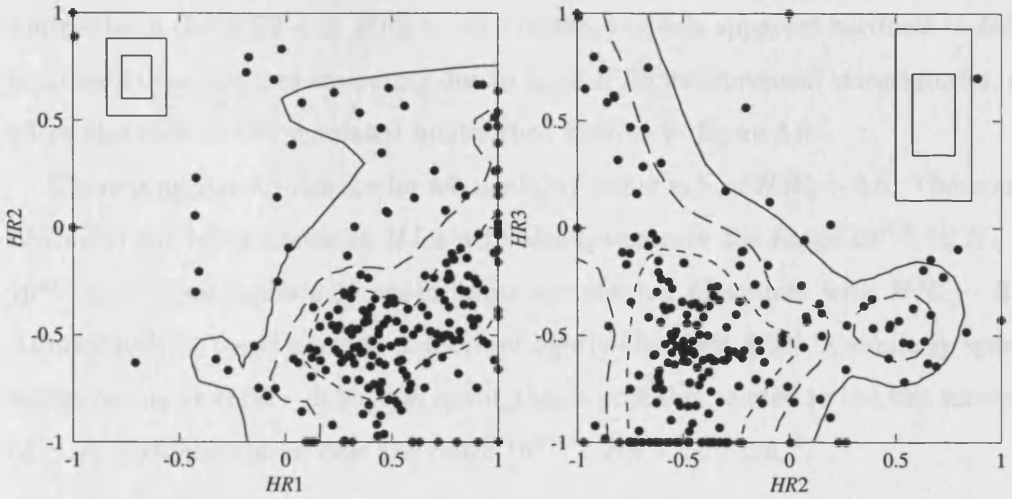

Figure 3.11: X-ray colour-colour distributions found in the 13^H sample. The levels of the contours are set such that they include 50% (short dash), 75% (long dash), and 90% (solid line) of the smoothed source distribution, and were generated in the same way as for figure 3.9. Typical sizes of σ_{HR1}, σ_{HR2}, and σ_{HR3} are shown with boxes, for sources having "two-band" fluxes of $10^{-14.5}$ and 10^{-14} erg s^{-1} cm^{-2}, where the "two-band" flux is the flux measured over the two energy bands used to calculate the hardness ratio.

3.6 Results

3.6.1 Colour distribution of the 13^H sample

Figure 3.10 also shows the colour-colour distributions of the 13^H sample, with grey-scale and contours generated in the same way as for figure 3.9. Figure 3.11 shows the same contours, but with the individual 13^H field sources overlaid.

It can be seen from these figures that there is a strong concentration of 13^H sources in the $(HR1, HR2, HR3) \simeq (0.4, -0.5, -0.5)$ region, approximately the location that would be expected for an unabsorbed AGN with a power law slope of $\Gamma \sim 1.9$. However, a large number of the 13^H sources have much higher values of $HR1$ and $HR2$ than the those typical for an unabsorbed AGN, indicating that absorption is present in a significant fraction of the sample. A number of the 13^H

sources lie in the $HR2 < 0$, $HR3 > -0.3$ region, but this apparent hardness in $HR3$ is probably the result of scattering due to large $HR3$ measurement uncertainties, an effect also seen in the simulated unabsorbed sources in figure 3.9.

There is apparent evidence for a bimodality either side of $HR1 = 0.6$. This could be due to the fast increase in $HR1$ with absorption over the range $10^{21.5} < N_H < 10^{22.5}$ cm^{-2}, (see figure 3.9) which limits the number of sources with $HR1 \sim 0.6$. Alternatively, it could point to a deficit of lightly absorbed AGN. A similarly sparse region occurs at $HR2 \sim 0.25$, and again, this is probably related to the fast increase of $HR2$ with absorption over the range $10^{23} < N_H < 10^{24}$ cm^{-2}.

3.6.2 Reproducing the source counts in the 13^{H} sample

Figure 3.12 compares the 0.5–2 keV band integral source counts ($N(> S_{0.5-2})$, normalised to the Euclidean slope) measured in the 13^{H} sample, with those predicted by the simulated model distributions. Because the 13^{H} sample and the simulated populations have an identical survey-depth/sky area relation, no correction is needed for sky coverage. There is a large disparity between the 13^{H} and simulated populations, especially around the knee of the observed $N(> S_{0.5-2})$ at $S_{0.5-2} \sim 10^{-14}$ erg s^{-1} cm^{-2}. The $f(N_H)$ models all predict rather similar $N(> S_{0.5-2})$ curves, especially at faint fluxes, where the statistical errors are smaller. This implies that the disparity between the sample and the models is primarily caused by differences between the 13^{H} sample and the Miyaji et al. (2000) XLF (and/or its evolution).

The primary purpose of this study is to compare the predictions of the $f(N_H)$ models, and so it is important to minimise the effect on the statistical analysis caused by the disparity between the 13^{H} sample and the XLF/evolution model. Therefore, I have examined the X-ray colour distributions, rather than the distributions of absolute fluxes. One would expect the colour-colour distributions to be more sensitive to $f(N_H)$ than to the form of the XLF; a small change of the position of an absorbed AGN in the z-L_X plane has a strong effect on observed flux, but only a small effect

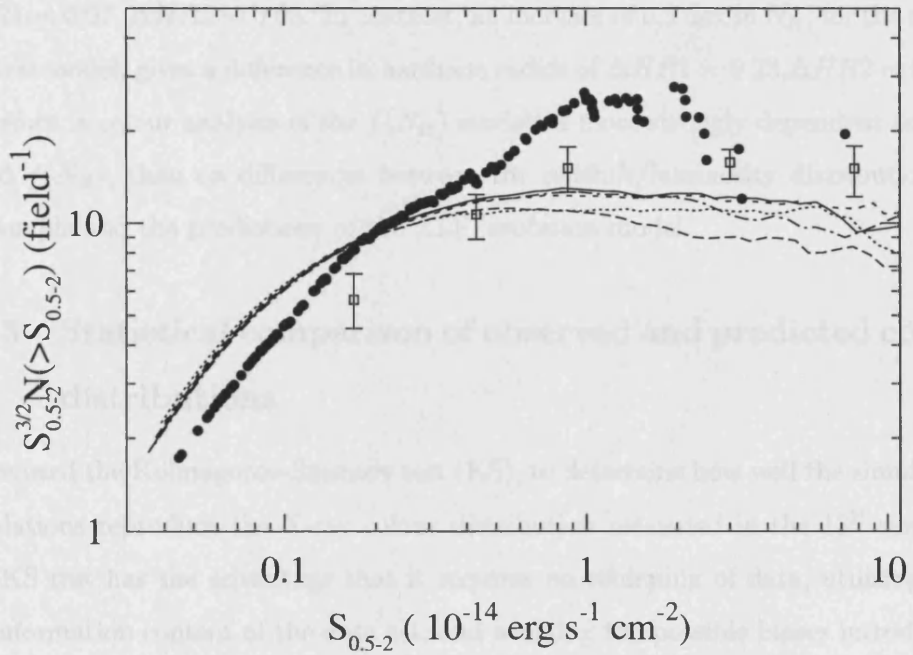

Figure 3.12: The $N(> S_{0.5-2})$ curves measured in the 13^{H} sample (filled circles), compared to those produced by the $(\log N_H)^2$ (solid line), $(\log N_H)^8$ (long dash), *GSH01A* (short dash), *GSH01B* (dotted), and *U03* (dot-dash), N_H distribution models. These results are for the absorbed power-law spectral model and are normalised to the Euclidean-slope. The data points from the sample of Miyaji et al. 2000 (taken from figure 6 of their paper), are also shown (open symbols with errorbars), and have been normalised assuming a sky area of 0.185 deg^2. The disparities between the observed source counts and the model predictions are discussed in section 3.8.4.

on its X-ray colours. For example, consider that the space density of AGN peaks at $z = 1.3$, rather than at $z = 1.6$ as predicted by the Miyaji et al. 2000 LDDE1 XLF model. The difference in hardness ratios between AGN at these two redshifts, each having an absorbed power-law spectrum ($N_H = 10^{22}$ cm^{-2}, $\Gamma = 1.9$), is $\Delta HR1 = 0.07$, $\Delta HR2 = 0.03$. In contrast, an increase of 0.5 dex in N_H, for the same spectral model, gives a difference in hardness radios of $\Delta HR1 = 0.23$,$\Delta HR2 = 0.17$. Therefore, a colour analysis of the $f(N_H)$ models is more strongly dependent on the tested $f(N_H)$, than on differences between the redshift/luminosity distribution of the sample and the predictions of the XLF/evolution model.

3.6.3 Statistical comparison of observed and predicted colour distributions

I have used the Kolmogorov-Smirnov test (KS), to determine how well the simulated populations reproduce the X-ray colour distribution measured in the 13^{H} sample. The KS test has the advantage that it requires no rebinning of data, utilising the full information content of the data set, and avoiding the possible biases introduced by the choice of bins. However, the KS test does not take into account the relative errors on data points, meaning that low signal to noise measurements can, to some extent, "wash out" the signal from the more precise measurements. I have used a three dimensional extension of the KS test (3D-KS), as devised by Fasano & Franceschini (1987), to compare the 13^{H} sample with the output simulated populations in ($HR1, HR2, HR3$) space. In order to examine more closely how the models reproduce the sample distribution, I have also carried out one dimensional KS tests separately on $HR1$, $HR2$ and $HR3$. Table 3.2 shows P_{3D-KS}, for the eight $f(N_H)$ models, and for both AGN spectral models that have been tested. The 3D-KS test probabilities have been calculated as described in section 3.6.4. These results of the KS tests, and the AGN population models that they favour are discussed in section 3.8.1.

Table 3.2: Results of Kolmogorov-Smirnov and three dimensional Kolmogorov-Smirnov tests. The probabilities (P_{3D-KS} and P_{KS}) were calculated by comparing the distributions of $HR1$, $HR2$ and $HR3$ predicted by the eight $f(N_H)$ models with the distributions found in the 13^H sample. Results are shown for both the absorbed power-law (APL) spectral model, and the absorbed power-law with reflection (APL+R) spectral model.

	$P_{3D-KS}(HR1, HR2, HR3)$		$P_{KS}(HR1)$		$P_{KS}(HR2)$		$P_{KS}(HR3)$	
N_H Model	APL	APL+R	APL	APL+R	APL	APL+R	APL	APL+R
$R = 0$	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$	2.6×10^{-4}	0.0027
$(\log N_H)^2$	$< 1 \times 10^{-5}$	1×10^{-5}	3.6×10^{-7}	3.1×10^{-6}	1.4×10^{-4}	0.0016	0.0052	0.034
$(\log N_H)^5$	0.0021	0.0019	2.5×10^{-4}	3.9×10^{-4}	0.0050	0.057	0.0078	0.039
$(\log N_H)^8$	0.0074	0.058	0.019	0.041	0.20	0.77	0.013	0.094
GSH01A	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	1.3×10^{-8}	4.5×10^{-8}	0.081	0.37	0.014	0.043
GSH01-R8	0.0014	0.0033	0.012	0.015	0.96	0.62	0.025	0.12
GSH01B	3.4×10^{-4}	0.0016	0.0058	0.0019	0.88	0.39	0.028	0.19
U03	$< 1 \times 10^{-5}$	1×10^{-5}	1.5×10^{-6}	4.2×10^{-6}	4.3×10^{-4}	0.0033	0.0038	0.026

3.6.4 Calculating 3D-KS test probabilities

As shown by Fasano & Franceschini (1987), the conversion from the 3D-KS test statistic, D_{3D-KS}, to the probability that two samples were taken from the same underlying population, P_{3D-KS}, is strongly dependent on the number of sources in the tested samples, and the degree of correlation between the tested variables. Fasano & Franceschini (1987) numerically generated lookup tables to allow this conversion at a number of confidence levels and values of the correlation parameter, and for a range of sample sizes. However, these tables give only a relatively small number of conversion values, at discrete confidence levels, sample sizes, and values of the correlation parameter. Therefore, I have run a set of Monte Carlo simulations, which permit conversion from D_{3D-KS} directly into P_{3D-KS} using the precise sample size and correlations parameters measured in the 13^{H} sample. I calculated a "three-dimensional probability density map" (3D-PDM) of the 13^{H} sample in $(HR1, HR2, HR3)$ space. The contribution from each source to the 3D-PDM is calculated from a 3D-Gaussian that has widths equivalent to σ_{HR1}, σ_{HR2}, and σ_{HR3}. The normalisation of the 3D-Gaussian for each source is set such that the contribution to the 3D-PDM is unity. The 3D-PDM is then used to generate sets of random populations, having 217 and 25000 members respectively, for which D_{3D-KS} is calculated. The set of 217 sources matches the size of the 13^{H} sample, and the set of 25000 is the approximate size of the output population for each N_H model. 100000 repetitions of the simulations were carried out. The 3D-KS probability for any particular value of D_{3D-KS}, is then equal to the fraction of the 100000 repetitions which have a D_{3D-KS} greater than this value. The absolute lower limit at which the probability can be estimated is given by the reciprocal of the number of simulation iterations, i.e. 0.001% for 100000 repetitions, although the uncertainty is large at this low probability level. The limit on the number of repetitions was determined by the processing time available.

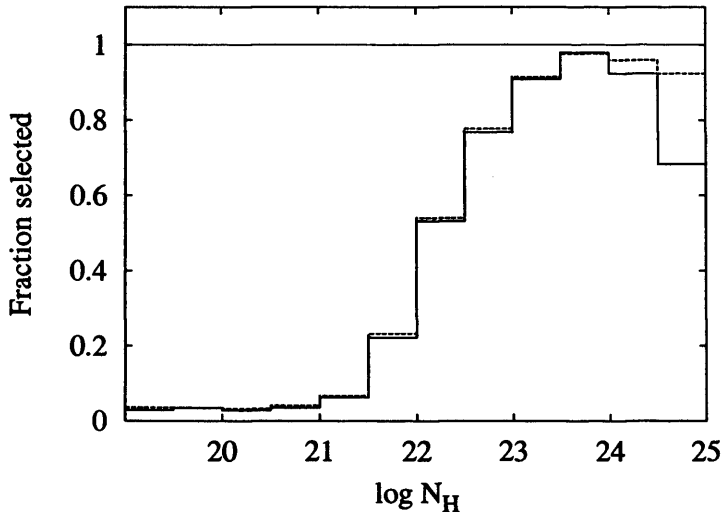

Figure 3.13: The fraction of output simulated sources which are selected by the hardness ratio cuts $HR1 - \sigma_{HR1} > 0.6$ and/or $HR2 - \sigma_{HR2} > -0.3$, as a function of absorption. The solid line shows the result for the absorbed power-law spectral model, and the dashed line shows the result for the absorbed power-law plus reflection spectral model.

3.7 Selecting absorbed sources from their X-ray colours

If for example, one wishes to examine in detail moderate to heavily absorbed AGN, then some X-ray colour selection criterion is required which will allow only this population to be chosen. Examination of the outputs from the simulated populations (see figure 3.9), shows that a cut of $HR2 > -0.3$ will select the majority of the most heavily absorbed sources ($N_H > 10^{22.5}$ cm^{-2}). A $HR2 > -0.3$ cut was shown to discriminate efficiently between optical type-1 and type-2 AGN in the *XMM-Newton* Bright Serendipitous Survey (XBSS) sample (Caccianiga et al., 2004; Della Ceca et al., 2004). However, it should be cautioned that the XBSS definition of $HR2$ differs slightly from that used here; they use the 2–4.5 keV energy band (rather than the 2–5 keV band used here), and report $HR2$ only for the MOS2 dataset. AGN with intermediate absorption (down to $N_H \sim 10^{21.5}$ cm^{-2}) can also be selected with the

condition $HR1 > 0.6$. In order to reduce the number of faint soft sources with low signal-to-noise HR measurements that are scattered into the "hard" sample, I require that a source must satisfy the HR conditions by more than 1σ to be included. That is, $HR2 - \sigma_{HR2} > -0.3$ and/or $HR1 - \sigma_{HR1} > 0.6$. Figure 3.13 demonstrates the effectiveness of these hardness ratio criteria in selecting only the simulated sources which have significant absorption. The slight dip in the selected fraction at $N_H > 10^{24}$ cm^{-2} is caused by the generally large errors on $HR1/HR2$ for the most heavily absorbed sources. This evaluation of the effectiveness of this absorbed source selection scheme assumes a simple absorbed power-law AGN spectral model. Spectral features, such as an additional soft component, will serve to degrade the efficiency. I have not used an $HR3$ criteria to select absorbed sources, because of the relatively poor average measurement accuracy for $HR3$, and its sensitivity to only very high columns. My "hard-source" selection criteria (based on $HR1$ and $HR2$), when applied to the 13^{H} sample, results in 86/217 "hard" sources ($40 \pm 3\%$ of the total). This fraction is consistent with the fraction ($34 \pm 9\%$) of *optical* type-2 AGN identified in the 2–4.5 keV selected subset of the XBSS (Della Ceca et al., 2004). The "hard" fractions predicted by each of the $f(N_H)$ models are presented in table 3.3, and are calculated by applying the selection criteria to the output simulated populations. For the absorbed powerlaw spectral model, the fractions of hard sources predicted by the $(\log N_H)^8$, *GSH01-R8*, and *GSH01B* models are consistent within 3σ with the fraction seen in the 13^{H} sample (0.40). By including a reflection component to the spectral model, the hard fraction is increased by less than 2% for each of the model N_H distributions. In the latter case, the hard fraction predicted by the *GSH01A* distribution is also consistent (at the 3σ level) with the hard fraction in the 13^{H} sample.

Table 3.3: The mean fractions of "hard" sources, h, (satisfying $HR1 - \sigma_{HR1} > 0.6$ and/or $HR2 - \sigma_{HR2} > -0.3$), predicted by each simulated $f(N_H)$ model. The corresponding h measured in the 13^{H} sample is 0.40. Results are shown for simulated populations generated using the absorbed power-law spectral model (APL), and the absorbed power-law plus reflection spectral model (APL+R). As a measure of the random scatter of the simulated populations, I show the standard deviation, σ_h, of the hard fraction over the 100 simulation repetitions.

	APL		APL+R	
$f(N_H)$ model	h	σ_h	h	σ_h
$R = 0$	0.026	0.010	0.029	0.010
$(\log N_H)^2$	0.226	0.024	0.239	0.026
$(\log N_H)^5$	0.278	0.025	0.288	0.024
$(\log N_H)^8$	0.332	0.025	0.347	0.029
GSH01A	0.300	0.023	0.313	0.029
GSH01-R8	0.399	0.024	0.413	0.028
GSH01B	0.420	0.022	0.436	0.028
U03	0.235	0.024	0.240	0.026

3.8 Discussion of results

3.8.1 Reproducing the X-ray colours in the 13^{H} sample

Table 3.2, shows that, when an absorbed power-law spectral model is used, none of the $f(N_H)$ models provide a good description of the X-ray colours in the 13^{H} sample; all the models are strongly rejected by the 3D-KS test (with greater than 99% confidence). The $(\log N_H)^8$ model provides the best fit with a probability of 0.7%. However, the mismatch must be partly due to the effects of the disparity between the data and the Miyaji et al. (2000) XLF/evolution model. The addition of a reflection component to the AGN spectra improves the P_{3D-KS} for almost all of the $f(N_H)$ models. The best fitting distribution is still the $(\log N_H)^8$ model, but it now has a much improved probability of 6%. The remainder of the $f(N_H)$ models are strongly rejected by the 3D-KS test, with greater than 99.5% confidence. Considering that the only tuned parameter is the overall normalisation of the XLF, the match between the $(\log N_H)^8$ model and the sample, is actually rather a good one. The large range of 3D-KS probabilities measured for the range of $f(N_H)$ models demonstrates that an X-ray colour analysis of *XMM-Newton* sources is indeed sensitive to the AGN absorption.

The results of the KS tests on individual hardness ratios reveal more clearly where the $f(N_H)$ models succeed or fail to reproduce the X-ray colours of the 13^{H} sources. The addition of a reflection component to the absorbed power-law spectral model improves the match in $HR3$ for all of the $f(N_H)$ models, and improves the match in $HR1$ and $HR2$ for the majority of the $f(N_H)$ models. When a reflection component is included in the spectral model, the $HR2$ distributions predicted by the $(\log N_H)^8$, *GSH01A*, *GSH01-R8* and *GSH01B* models closely match ($P_{KS} > 30\%$) the $HR2$ distribution of the 13^{H} sample. A far greater difference between the predictions of the models and the 13^{H} sample is evident in the $HR1$ and $HR3$ distributions. In comparison to the 13^{H} sample, all of the $f(N_H)$ models predict an overly large fraction of sources having very hard colours at energies below 2 keV (i.e. $HR1 =$

1). In the 13^{H} sample, this fraction is only 8%; however, even in the best fitting $(\log N_H)^8$ model, the fraction of sources with $HR1 = 1$ is 15%. Examination of the simulated populations reveals that almost all (90%) of the sources with $HR1 = 1$ are significantly absorbed ($N_H > 10^{22}$ cm^{-2}), and so have had virtually all their flux below 0.5 keV removed. However, these hard simulated sources have similar distributions in L_X, z and Γ space to the rest of the simulated populations. The relatively low value of $P_{KS}(HR1)$ for the *GSH01A*, *GSH01-R8* and *GSH01B* models can be partly attributed to the low numbers of AGN that they predict to have light absorption (in the range $10^{20} N_H < 10^{22}$ cm^{-2}). Within this subset of models, the evolving *GSH01B* model is preferred over the *GSH01A* model, but is marginally less successful than the *GSH01-R8* model. However, the latter model is somewhat unphysical in that it contains a much greater ratio of absorbed to unabsorbed sources than is seen in the local universe (Risaliti, Maiolino & Salvati, 1999).

The $f(N_H)$ models produce too large a fraction of simulated sources having $HR3 = -1$. This is most probably due to the over-abundance of very faint sources predicted by the model populations (see figure 3.12). These faint sources are detected just above the flux limit in the softer bands, but have count rates which fall below the background level in the hardest band, and hence are measured to have $HR3 \approx -1$.

The statistical analysis strongly rejects the *U03* model in all tests. This is in agreement with the findings of the study by Treister et al. (2004), which is based on deep multi-wavelength data in the GOODS fields. These authors tested the $f(N_H)$ model of Ueda et al. (2003) (on which the *U03* model is based), alongside a simpler, non-evolving $f(N_H)$, but found that the latter provided a much better description of the data.

3.8.2 The X-ray colours of the "hard" sources

By examining the subset of sources satisfying the "hard" selection criteria (see section 3.7), a comparison can be made between the observed and predicted distributions of absorption in the subset of sources with $N_H > 10^{22}$ cm^{-2}. To make this comparison, I have carried out 3D-KS and KS tests on $HR1$, $HR2$, and $HR3$, as before, but just for the "hard" selected subsets of the 13^{H} sample and the simulated populations. The 3D-KS test rejects each of the $f(N_H)$ models with high confidence, (both with and without a reflection component included in the model spectra). I have examined the one-dimensional KS test results to determine the source of this large disparity. I find that the KS probabilities for the best fitting $(\log N_H)^8$ model (with the absorbed power-law spectral model), are 0.0003, 0.76, and 0.12, for $HR1$, $HR2$ and $HR3$ respectively. The equivalent probabilities when an additional reflection component is included in the spectral model are 0.0002, 0.77, and 0.29. In fact, the $HR2$ and $HR3$ distributions predicted by all of the $f(N_H)$ models (excepting the $R = 0$ model) provide rather good matches to the $HR2$ and $HR3$ distributions found in the "hard" subset of the 13^{H} sample. This is probably because of the rapid decline in the selected fraction of "input" sources for high absorbing columns (see figure 3.8), which acts to diminish the importance of the differences in the shape of the $f(N_H)$ models above $N_H = 10^{22}$ cm^{-2}. The addition of a reflection component to the spectral model improves the KS probability for $HR3$ by a factor of ~ 2 for each of the $f(N_H)$ models.

It is evident that the differences between the observed and predicted $HR1$ distributions are more pronounced in the "hard" subset, than the in the whole population. This appears to be caused by the $f(N_H)$ models predicting too many output sources having $HR1 = 1$, an effect that is more pronounced in the "hard" sub-sample. The fraction of the "hard" sub-sample with $HR1 = 1$ is 20% for the 13^{H} field, but is predicted to be $\sim 40\%$ for the $f(N_H)$ model populations. The disparity could be

explained if a number of the heavily absorbed AGN in the 13^{H} sample have an additional soft X-ray component in their spectra. In order to reproduce the distribution of $HR1$, this phenomenon should occur in around 10–20% of the heavily absorbed sources in the 13^{H} sample. A number of absorbed AGN with excess soft emission have been observed by other authors in samples of spectroscopically identified X-ray sources (e.g. Caccianiga et al., 2004; Mateos et al., 2005a; Page et al., 2006). The soft excess component could be due to intense starbursts in the host galaxy, or to diffuse emission surrounding an AGN embedded in a galaxy cluster. Alternatively, it could be scattered radiation from the central engine of the absorbed AGN.

3.8.3 Implications for AGN torus models

For the simplest toy model of a torus with uniform density, and a typical opening angle, θ_o, the fraction of AGN for which our line of sight passes through the torus is $cos(\theta_o)$. If the size of the "hard" fraction of the 13^{H} sample ($\sim$40%) is used as a measure of the number of absorbed AGN, it infers a rather wide torus opening angle of $\theta_o \sim 67°$. However, this estimate does not take into account the effect of the drop in the selection function toward high N_H, and should be seen only as an upper limit on θ_o. The selection function for the "hard" population relative to the selection function for the entire population can be estimated from the simulated AGN populations. Applying this correction to the 13^{H} sample, I predict an *intrinsic* "hard" fraction of $\sim$0.8, implying an opening angle of $\theta_o \simeq 37°$. If in the correction for the relative selection function, those sources with absorbing column above $N_H = 10^{24}$ cm^{-2} are excluded (where the 13^{H} sample constrains the models only weakly), then $\theta_o \simeq 52°$. Conversely, one can examine the range of torus parameters that would best match each of the $f(N_H)$ models. For the best fitting $(\log N_H)^8$ model, where the fraction of input sources with $N_H > 10^{22}$ cm^{-2} is $\sim$0.75, the predicted opening angle is $\theta_o = 42°$.

As demonstrated in figure 3.9, $HR1$ and $HR2$ are sensitive to the shape of

the distribution over a wide range of N_H, particularly for intermediately absorbed sources. As illustrated in table 3.2, the $(\log N_H)^8$ N_H model is a much better match to the $HR1$ distribution of sources in the 13^{H} than the *GSH01A* model. These two N_H models are very similar in the range $10^{22} < N_H < 10^{24}$ cm^{-2}, they predict a similar fraction of unabsorbed AGN ($N_H < 10^{21}$ cm^{-2}), and predict comparable numbers of "hard" sources. Therefore, the large difference in the KS test results for these two models must lie primarily in the $10^{21} < N_H < 10^{22}$ cm^{-2} range, in which the $(\log N_H)^8$ model contains many more AGN. A major problem with any "unified" scheme in which all AGN have similar, uniformly dense tori, is that nearly all AGN will be either heavily absorbed or completely unabsorbed. However, more complex models, incorporating a wide distribution of torus densities, do predict larger numbers of intermediately absorbed AGN. For example, a model in which the density falls off exponentially with angle away from the plane of the torus, predicts a much flatter $f(N_H)$ (e.g. Treister et al., 2004). It is possible, with some tuning of such a model's parameters, to approximately match the best fitting $(\log N_H)^8$ model distribution.

Because absorbing columns in the $10^{21} < N_H < 10^{22}$ cm^{-2} range have a noticeable impact only on $HR1$, such an effect would not have been detectable in the X-ray colour distributions if the 0.2–0.5 keV band had not been considered. A number of studies of the absorption distribution in faint AGN have based their estimates of N_H on hardness ratios between the 0.5–2 and 2–10 keV bands, and therefore may have underestimated the number of intermediately absorbed AGN (e.g. Ueda et al., 2003; Treister et al., 2004).

3.8.4 Source count disparity

Each of the $f(N_H)$ models predict similar 0.5–2.0 keV integral source count-flux relations, $N(> S_{0.5-2})$. However, these are seen to reproduce poorly the $N(> S_{0.5-2})$ relation that is observed in the 13^{H} sample (see figure 3.12). The simulated populations

under-produce $N(> S_{0.5-2})$ above the normalisation flux (2×10^{-15} erg s^{-1} cm^{-2}), and over-produce $N(> S_{0.5-2})$ below this flux. In fact, at 10^{-14} erg s^{-1} cm^{-2} the simulated populations under-produce the source counts seen in the 13^{H} sample by a factor of about two. This disparity is seen to a similar degree in each of the $f(N_H)$ models, suggesting that it is related primarily to the difference between the redshift/luminosity distribution of the 13^{H} sources and the predictions of the Miyaji et al. (2000) LDDE1 XLF model. The $N(> S_{0.5-2})$ of the Miyaji et al. (2000) sample is also shown in figure 3.12, plotted assuming a sky coverage of 0.185 deg^2. This comparison illustrates that in the flux range $10^{-14} - 10^{-13}$ erg s^{-1} cm^{-2}, the LDDE1 XLF model under-produces the source counts of the sample from which it was derived. The shape of the $N(> S_{0.5-2})$ relation of the Miyaji et al. (2000) sample is closer to that seen in the 13^{H} sample than to the $N(> S_{0.5-2})$ predicted by the models. The faintest AGN in the Miyaji et al. (2000) sample are from the deepest *ROSAT* observations of the Lockman Hole field, where the flux limit of the data was $\sim 2 \times 10^{-15}$ erg s^{-1} cm^{-2}. The deeper flux limit in the 13^{H} data explores a part of $L - z$ space outside that constrained by the sample of Miyaji et al. (2000). A previous comparison of source counts from *ROSAT* observations in the Lockman Hole and 13^{H} fields, revealed a $\sim 10 - 20\%$ over-abundance near $S_{0.5-2} = 10^{-14}$ erg s^{-1} cm^{-2} in the 13^{H} field with respect to the Lockman Hole (McHardy et al., 1998). In addition, Loaring et al. (2005) found that the 13^{H} field is slightly overdense in the 0.5–2 keV band with respect to both of the Chandra deep fields North and South. The differences between the model predictions and the 13^{H} sample are probably caused by a combination of these factors. In particular, the poor match to the 13^{H} source counts provided by the Miyaji et al. (2000) LDDE1 XLF model suggests that their complex scheme may require some revision.

3.8.5 High redshift AGN in the 13^{H} field

The shape of the XLF at high redshift is poorly known because of the difficulties in obtaining a large spectroscopically identified sample of these objects. The simulated AGN populations can be used to predict the number of high-z AGN that are expected in the 13^{H} sample. About 16% of the output sources in the simulated populations lie at $3 < z < 5$. Therefore, one should expect around 35 of the 13^{H} X-ray detections to lie at these redshifts. However, only a single AGN has been identified at $z > 3$ by the follow up optical spectroscopy program in the 13^{H} field (which has secure identifications for over 100 sources). The disparity could be related to the over-production of faint sources by the Miyaji et al. (2000) XLF model; these are more likely to be at high redshift. The X-ray detection probability of AGN at $z > 3$ is much less dependent on N_H than for AGN at lower z; unabsorbed hard X-rays are redshifted to energies where EPIC is most sensitive. Therefore, most of the $f(N_H)$ models predict that absorbed AGN make up the majority of the *detected* high-z population, the precise fraction being dependent on the particular $f(N_H)$ model. However, the absorption of optical and UV spectral features does severely affect the probability of identification for high redshift, absorbed AGN. A definitive conclusion will not be possible until the identification program reaches a higher level of completeness. Further spectroscopic programs on 8m class telescopes, and redshift determination using photometric techniques should solve this problem.

3.9 Summary

I have demonstrated how a colour-based analysis of deep *XMM-Newton* observations can be used to constrain models of absorption in the AGN population without requiring complete optical spectroscopic follow up. By using a detailed simulation technique, it is possible to take account of the complex selection function at work in the sample, and understand how this modulates the input population. A simple $f(N_H)$ model together with an absorbed power-law spectral model (including a

reflection component), reproduces the $HR1, HR2, HR3$ colour distribution of the 13^{H} sources with probability 6%. The other model N_H distributions were all rejected at greater than 99.5% probablity. In particular, two more complex $f(N_H)$ models are strongly rejected by the 3D-KS test; the redshift dependent *GSH01B* model produces too many hard sources, and the luminosity dependent *U03* model produces too few. In general, the addition of a reflection component to the absorbed power-law spectral model improved the match between the colour distributions of the models and the sample. The reflection component serves to harden the spectral slope at higher energies, and its effect was most evident in the $HR2$ and $HR3$ distributions.

There is a large disparity between the shape of the $N(> S_{0.5-2})$ predicted by the model populations and that found in the 13^{H} sample. I suggest that this is for the most part due to differences between the luminosity/redshift distribution of AGN in the 13^{H} field, and the predictions of the XLF/evolution model used to represent the model populations. The differences will have had some impact on the colour distributions predicted by the $f(N_H)$ models, and could explain the surfeit of $HR3 = -1$ sources in the simulated populations.

There is some evidence that the spectra of a significant fraction of absorbed sources in the 13^{H} sample have an additional soft X-ray component. This feature was not included in the simple AGN spectral models, and therefore may have caused the disparity between the $HR1$ distributions predicted by the models and the $HR1$ distribution of the 13^{H} sources. Considering these factors, the $(\log N_H)^8$ model, which is acceptable at the 6% level, provides a reasonably good match to the X-ray colours of the 13^{H} sources.

3.10 Extending the analysis to other fields

I have shown that an X-ray colour analysis of sources detected in the *XMM-Newton* observations of the 13^{H} field is able to differentiate between a number of N_H distribution models. However, the AGN population models that I have explored are not

able to reproduce fully the X-ray properties of the 13^{H} sources. In order to improve the constraints on the AGN population models, one must study the X-ray properties of a sample of X-ray sources for which the redshifts are known. The 500 ks of *XMM-Newton* observations in the *Chandra* Deep Field South (CDFS) are ideal for this, as shown in the next chapter. The addition of the redshift information provides a much better constraint on whether the N_H distribution evolves in luminosity and/or redshift space.

An alternative and complementary approach is to expand the study undertaken in this chapter to a larger sample of X-ray sources with similarly deep *XMM-Newton* coverage. This is what I do in chapter 5, expanding the size of the X-ray source sample by a factor of $\sim$8. I show that this larger sample also permits an investigation of how large a role is played by field-to-field variations.

Chapter 4

The AGN population in the CDFS: the *XMM-Newton* view

4.1 Introduction

My study in chapter 3 used 200 ks of (initially proprietary) *XMM-Newton* observations in the 13^{H} field. Since that time, a wealth of deep field data has become available in the *XMM-Newton* public archive. In this chapter I have taken advantage of the $\sim$ 500 ks of *XMM-Newton* EPIC observations which have been made of the *Chandra* Deep Field South (CDFS). In addition to the longer *XMM-Newton* exposure time in the CDFS relative to the 13^{H} field, this small patch of sky has been the target for a wealth of observations over many wavebands. Two additional data sets are particularly important for my work. The whole field has been covered by a mosaic of 4 $\times$ 250ks *Chandra* pointings, and the centre part covered with 1Ms of observations (Giacconi et al., 2002; Lehmer et al., 2005). As shown later, most of the *XMM-Newton* detections also appear in the *Chandra* catalogues. The high spatial precision of the positions determined with *Chandra* permits association of the *XMM-Newton* sources with unambiguous optical counterparts. The other key ingredient for this work is the large number of spectroscopic and/or photometric

redshift determinations which have been made in this field (Szokoly et al., 2004; Wolf et al., 2004; Zheng et al., 2004).

These additional data are an important aid to the determination of the absorption distribution of the AGN population. By including the redshift information the absorption and luminosity of individual *XMM-Newton* sources can be calculated. This allows a more thorough investigation of the redshift and/or luminosity dependence of the absorption distribution in the AGN population. The *XMM-Newton* observations of the CDFS (hereafter XMM-CDFS) provide a superb dataset for measuring the spectral properties of faint X-ray sources. The EPIC imaging reaches to fluxes well below the break in the 2–5 keV source counts, covers around 0.19 deg^2 and contains enough photons to permit broad band X-ray spectral analysis for even the faintest sources. The studies of Streblyanska et al. (2004) and Braito et al. (2005) have also used the XMM-CDFS dataset to investigate the X-ray spectra of a number of the brighter AGN in the field. However, until this work, there has been no investigation of the X-ray properties of the entire source population detected in the *XMM-Newton* imaging.

This chapter is laid out as follows. In section 4.2.1 I describe the reduction of the *XMM-Newton* data, and the source detection process. In 4.3 I detail how I have correlated this with the other datasets available in the CDFS. In section 4.5 I describe a novel Monte Carlo method for calculating the absorption and luminosity of the sample, and demonstrate its fidelity. In section 4.6 I present the distribution of absorption, luminosity and redshift in the XMM-CDFS sample and compare it to the predictions of a number of AGN population models.

In this chapter L_{2-10} refers to an object's *intrinsic* X-ray luminosity (that is, before absorption), in the *rest-frame* 2–10 keV energy band. The majority of the work presented in this chapter has now been published in Dwelly & Page (2006).

4.2 Observations and data reduction

4.2.1 The *XMM-Newton* dataset

The XMM-CDFS data consist of two observations carried out in July 2001, and six in January 2002. The observations all have similar pointing centres (approximately $RA = 03^h32^m27^s$, $Dec = -27°48'55''$), but the Jan 2002 observations have position angle rotated $\sim 180°$ with respect to the July 2001 observations. A summary of the XMM-CDFS observations is given in table 5.1. The XMM-CDFS data cover ~ 0.19 deg^2 (nearly twice the sky area of the 1Ms *Chandra* observations), and total around 500 ks. All three EPIC detectors (MOS1, MOS2 and pn) were operated with the 'Thin1' filters and were in full frame mode. After temporally filtering periods of enhanced particle background from the event lists, there remain approximately 340 ks of *pn*, and 395 ks of MOS1/MOS2 exposure time. Images, exposure maps, and background maps were produced from the raw *XMM-Newton* data following the steps detailed in section 2.1; however, several additional considerations were made. In several observations I noticed an enhancement of the 0.2–0.5 keV background level for CCD #5 of MOS1. As a conservative measure I discarded all of the data from this chip in this energy range. This high background effect has been reported by Pradas & Kerp (2005). The astrometric positions of the XMM-CDFS data are tied to the positions of (relatively) bright point-like X-ray sources detected in the 1Ms *Chandra* imaging of the field, taking into account the (-1.1″, 0.8″) offset between the *Chandra* positions and optical counterparts (Giacconi et al., 2002).

Visual inspection of the 0.2–0.5 and 0.5–2 keV images reveals four regions of large scale ($> 1'$ diameter) diffuse emission, the locations of which are shown in figure 4.1. The most likely origin of this emission is from highly ionised gas in galaxy groups or clusters, the study of which is outside the scope of this thesis. The background fitting algorithm I have employed is not designed to remove diffuse emission on these scales. This unsubtracted diffuse emission will affect the measurement of the X-ray spectral properties of any AGN lying in these four regions. For these reasons I have

excluded these regions from the analysis. The total sky area removed amounts to ~ 28 arcmin^2 (4% of the total *XMM-Newton* sky coverage).

4.2.2 *XMM-Newton* source detection procedure

I have used the multi-band background fitting and source searching procedure described in section 2.1.8 to detect sources in the XMM-CDFS images. For the purposes of source detection (in which an off-axis dependent PSF model is used), the position of the optical axis in the combined images is set to be the *pn* exposure-weighted mean of the pointings of the eight separate observations. The four areas of extended emission are excluded by preventing the final iteration of EMLDETECT from placing sources in these regions. The sky density of sources is higher in the XMM-CDFS than the $13^{\rm H}$ field data because of the increased depth of the *XMM-Newton* observations. To account for this factor, the EMLDETECT source detection routine was carried out using a slightly modified set of parameters to those used in chapter 3. Namely, the closest distance permitted between adjacent sources was set to be the local "75% energy enclosed" radius of the EPIC PSF, whereas in chapter 3 the default "90% energy enclosed" radius was used. The most prominent areas of extended emission in the XMM-CDFS have been manually excluded, therefore it was possible to reduce computer processing time by running the source detection process over a smaller detection parameter space. Therefore I set the maximum FWHM permitted for an extended source to 5 pixels (20.5″, default 20 pixels), and the minimum improvement in detection likelihood for which a source would be considered to be extended to 15 (default 10).

As before, the vignetting corrected count-rates were converted to fluxes using XSPEC, assuming a power-law spectrum with photon-index 1.7, corrected for Galactic absorption of 8×10^{19} cm^{-2} (Rosati et al., 2002). The energy conversion factors for the CDFS field are listed in table 2.1.

At a minimum multi-band DET_ML level of 5.0, the raw XMM-CDFS sourcelist

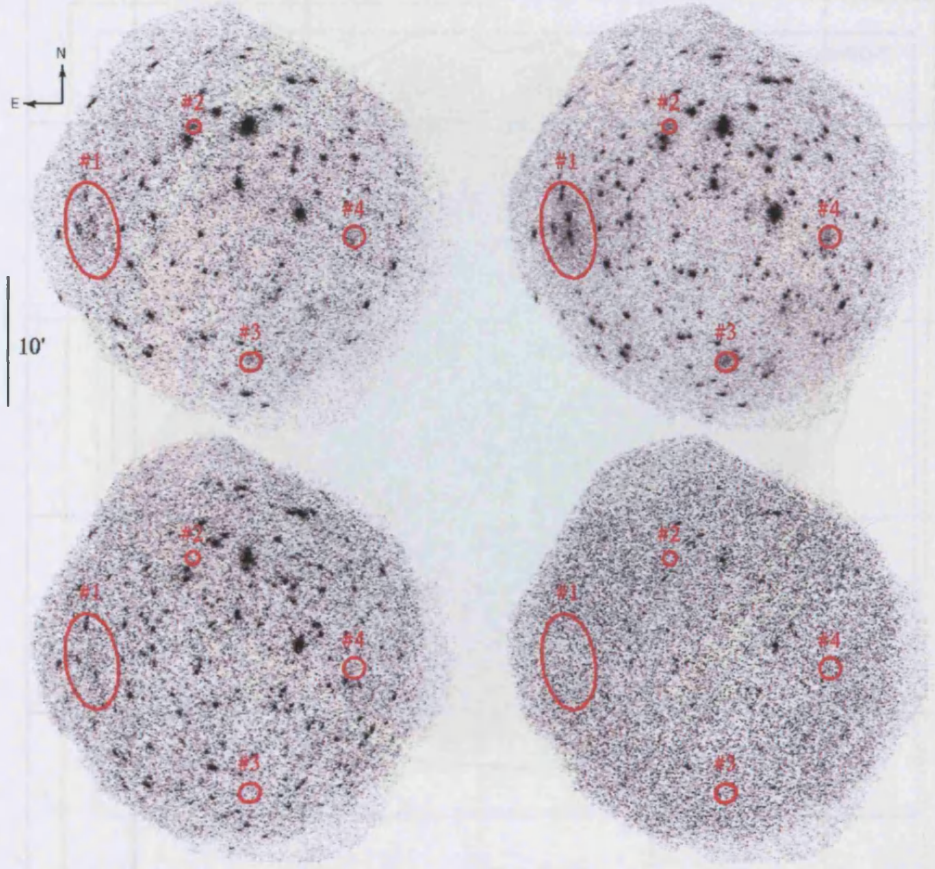

Figure 4.1: The combined XMM-CDFS EPIC MOS+pn images, background subtracted, and displayed on a square-root scale. The energy bands are 0.2–0.5 keV (top left), 0.5–2 keV (top right), 2–5 keV (bottom left) and 5–10 keV (bottom right). The field of view of EPIC is approximately 30′ in diameter. The four regions that were excluded due to the presence of extended emission are shown with numbered ellipses. These are likely to be galaxy clusters/groups and are discussed in section 4.8.

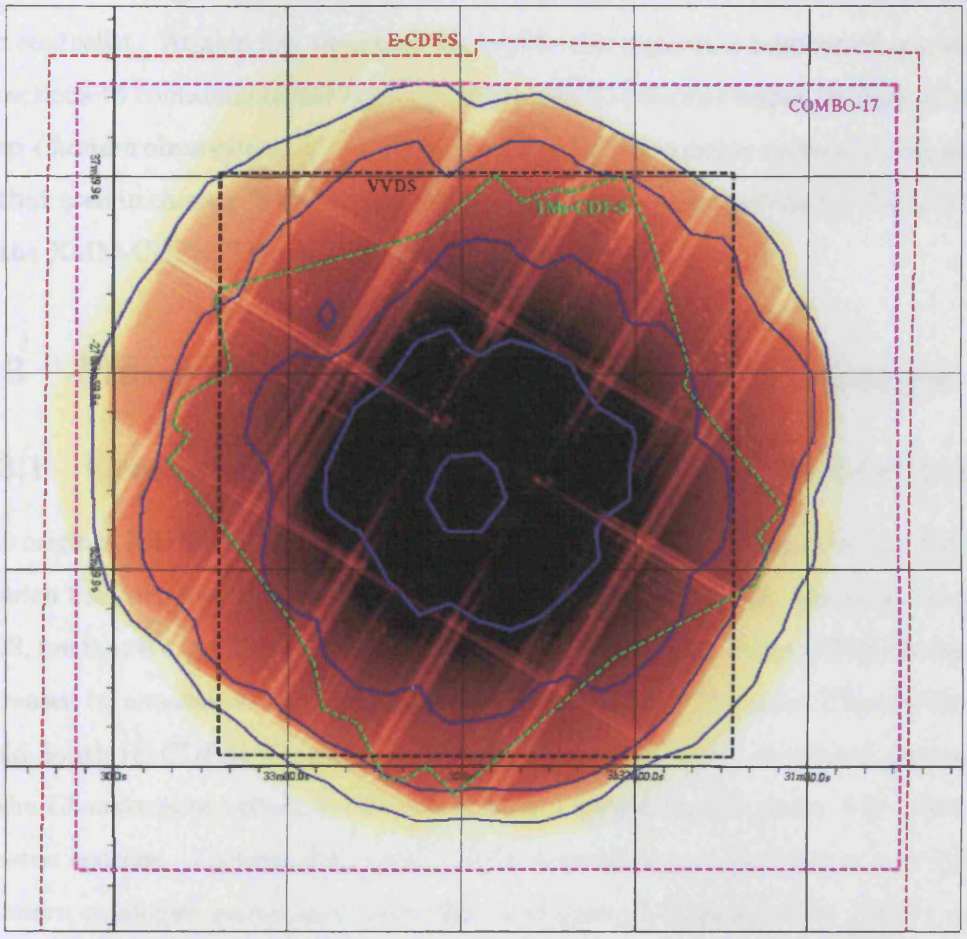

Figure 4.2: The combined *XMM-Newton* MOS1+MOS2+pn 0.5–2 keV exposure map. The peak pn-equivalent exposure time is ~540 ks. Contours (blue) are shown in 100 ks steps. I show the regions covered by the COMBO-17 survey, (large magenta dashed rectangle), the 1Ms *Chandra* imaging (green dashed polygon), and E-CDFS field (large red dashed polygon), and the approximate area covered by the VVDS (black dashed rectangle).

contains 435 sources. Four of these detections have very poorly determined positions ($\sigma_{pos} > 5''$), or have poorly determined extent (the 90% error on the measurement of the extent is greater than the extent itself), and so I have removed these from the sourcelist. At this low detection threshold, one expects a number of spurious detections to contaminate the faint end of the *XMM-Newton* sample. In light of the deep *Chandra* observations of the field, I have used an alternative method (compared to that used in chapter 3) to determine the minimum detection criteria for the sources in the XMM-CDFS. This is discussed in section 4.3.

4.3 Matching to *Chandra* and optical catalogues

4.3.1 Cross correlation with *Chandra* observations of the field

The original 1Ms *Chandra* imaging of the CDFS covers the central part of the *XMM-Newton* field of view (FOV) to great depth (Giacconi et al., 2002; Alexander et al., 2003, see figure 4.2). More recently, the *Chandra* sky coverage of the CDFS has been increased by a mosaic of four 250 ks *Chandra* pointings: the Extended *Chandra* Deep Field-South (E-CDFS) (Lehmer et al., 2005). I use the higher positional accuracy of the *Chandra* observations to aid unambiguous optical identification of the *XMM-Newton* sources. I match the XMM-CDFS detections to sources in a combined *Chandra* catalogue constructed from the catalogues of Giacconi et al. (2002) and Lehmer et al. (2005). As mentioned earlier, the (-1.1″, 0.8″) offset between the *Chandra* positions and optical counterparts in the Giacconi et al. (2002) catalogue is taken into account. For those *Chandra* sources which appear in both the Giacconi et al. (2002) and Lehmer et al. (2005) catalogues, I primarily used the positions from the former. The point spread function of the *XMM-Newton* EPIC detectors is strongly dependent on the off-axis angle (see section 2.2.1). At large off-axis angles the azimuthal component of the PSF becomes rather extended, whereas the radial component remains relatively constant. Therefore, I have matched the XMM-CDFS

detections to *Chandra* counterparts using an ellipsoidal region. The semi-major axis of this ellipse is increased from 5″ for sources at the centre of the FOV, up to a maximum of 10″ for sources at off axis angles greater than 15′. The semi-minor axis is kept constant at 5″, and is oriented parallel to the line joining the source position and the nominal optical axis of the EPIC-pn detector. In addition, for the few XMM-CDFS detections which EMLDETECT determines to be slightly extended, both semi-axes of the search ellipse are increased by the measured extent. This choice of X-ray position matching criteria is discussed further in section 4.5.6.

Using these positional criteria, 330 of the 431 XMM-CDFS detections are matched to *Chandra* sources; 185 of these matches are to sources in the 1Ms *Chandra* catalogue, and 145 are to sources in the E-CDFS catalogue. For the XMM-CDFS sources with no *Chandra* counterpart, I have manually examined the *XMM-Newton* and *Chandra* images. In 15 cases, there is a nearby *Chandra* counterpart just outside the matching ellipse. I adopt the *Chandra* positions for these sources, and consider them to be adequate matches. Figure 4.3 shows the positional offsets between the matched XMM-CDFS and *Chandra* sources.

As discussed in chapter 3, the determination of the *XMM-Newton* detection likelihood limit is a balance between the desire to include as many sources in the XMM-CDFS sample as possible, against the need to minimise the number of spurious detections. One option would be to simply reject all *XMM-Newton* detections that are not matched to *Chandra* sources. However, whilst the headline flux limits achieved by the *Chandra* observations are fainter than those for the XMM-CDFS, the coverage is not uniform over the *XMM-Newton* FOV. What is more, the relative sensitivity of the *XMM-Newton* and *Chandra* detectors varies with energy. In particular, *XMM-Newton* EPIC is much more sensitive than *Chandra* at very high photon energies (> 5 keV), and at very low photon energies (< 0.5 keV), so sources having either very hard or very soft spectra will be preferentially detected with *XMM-Newton*. The *XMM-Newton* and *Chandra* observations of the CDFS

span approximately four years, therefore intrinsic variability on timescales of several months to years could also account for sources appearing in some catalogues and not others. For these reasons, and because this study is based primarily upon *XMM-Newton* data, I have curtailed the XMM-CDFS sourcelist purely on the basis of the *XMM-Newton* detection likelihood, DET_ML. However, the minimum level of DET_ML is set so that approximately 90% of the *XMM-Newton* detections have *Chandra* counterparts. At a detection likelihood threshold of 8.5, there are 335 *XMM-Newton* detections and 302 (90.1%) of these have at least one *Chandra* counterpart. This is the detection likelihood threshold used for the remainder of the chapter.

4.3.2 Source confusion in the XMM-CDFS sample

The deep *Chandra* coverage (with its narrow PSF) allows direct determination of the amount of source confusion in the XMM-CDFS images. There are 15 XMM-CDFS detections which have more than a single *Chandra* source inside the off-axis dependent matching ellipse. In order to determine whether these are genuinely confused sources, I have manually inspected the *XMM-Newton* images and the 1Ms/E-CDFS *Chandra* images provided by Alexander et al. (2003) and Lehmer et al. (2005). For one of these XMM-CDFS detections, the "confusion" appears to be the result of a single real astrophysical source appearing in both the 1Ms and E-CDFS catalogues. The E-CDFS source is the preferred counterpart in this case because the detection is at the edge of the 1Ms coverage. In the other fourteen cases, the *XMM-Newton* detection is matched to a clearly separated pair of *Chandra* sources. However, for five of these, there is a large brightness contrast (greater than factor of 5 in flux) between the two *Chandra* sources, and so I do not consider these XMM-CDFS sources to be "confused". The remaining 9 truly "confused" *XMM-Newton* detections are removed from the final XMM-CDFS sample because their *XMM-Newton* determined properties are superpositions of more than one real astrophysical source. The small

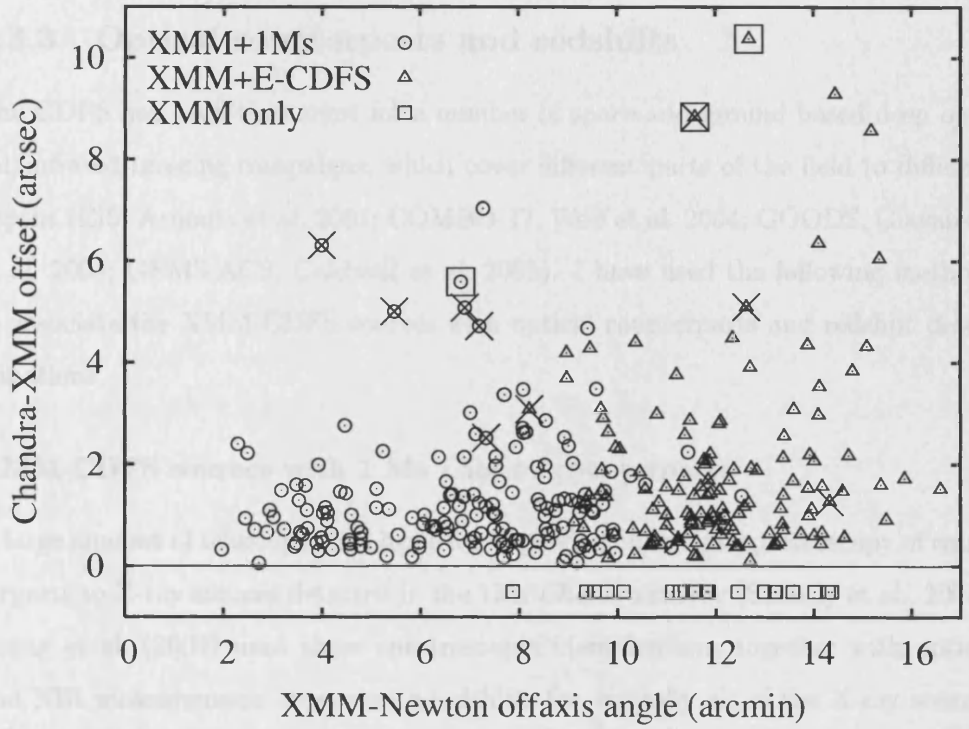

Figure 4.3: The differences between the positions determined using *XMM-Newton* and *Chandra* as a function of *XMM-Newton* off-axis angle. The *XMM-Newton* detections which have counterparts in the 1 Ms *Chandra* catalogue are shown with circles, those with an E-CDFS counterpart are shown with triangles. The *XMM-Newton* detections that are determined to be "confused" are marked with crosses, and the manual matches to *Chandra* sources are highlighted with large boxes. XMM-CDFS sources which do not have *Chandra* counterparts are shown at offset = -0.5.

number of detections that have to be removed demonstrates that source confusion plays only a small role ($\sim$ 3% of sources) in *XMM-Newton* surveys of high Galactic latitude fields, even in observations totalling several hundred kiloseconds.

4.3.3 Optical counterparts and redshifts

The CDFS has been the target for a number of space and ground based deep optical/infrared imaging campaigns, which cover different parts of the field to different depths (EIS, Arnouts et al. 2001; COMBO-17, Wolf et al. 2004; GOODS, Giavalisco et al. 2004; GEMS-ACS, Caldwell et al. 2006). I have used the following methods to associate the XMM-CDFS sources with optical counterparts and redshift determinations.

XMM-CDFS sources with 1 Ms *Chandra* counterparts

A large amount of telescope time has been expended on optical spectroscopy of counterparts to X-ray sources detected in the 1Ms *Chandra* survey (Szokoly et al., 2004). Zheng et al. (2004) used these spectroscopic identifications together with optical and NIR measurements to estimate redshifts for virtually all of the X-ray sources in the 1Ms *Chandra* catalogue of Giacconi et al. (2002). Initially, I adopted the Zheng et al. (2004) estimates for all the XMM-CDFS detections which are matched to *Chandra* sources in the 1Ms catalogue. However, as noted by Barger et al. (2005), a number of the Zheng et al. (2004) optical counterparts have relatively large offsets from the X-ray source positions. The X-ray–optical position offsets listed by Zheng et al. (2004) indicate that they have not applied the (0.8″,-1.1″) offset to the X-ray positions of Giacconi et al. (2002).

Therefore I have manually examined the *Chandra vs* optical positions for all of the CDFS sources where the optical position stated by Zheng et al. (2004) is more than 2″ from the 1Ms *Chandra* position, or where the position difference listed by Zheng et al. (2004) is greater than 2″, or where there is more than one COMBO-17

source within 2″ of the 1Ms *Chandra* position. For the other sources, the correct optical counterpart is likely to have been chosen by Zheng et al. (2004). I have drawn upon the *Chandra* 1Ms images[1] (Alexander et al., 2003), the E-CDFS images[2] (Lehmer et al., 2005), the COMBO-17 optical images[3] (Wolf et al., 2004), and the GEMS/GOODS ACS images[4] (Caldwell et al., 2006), to choose the most likely optical counterparts. For sixteen cases I decided that an alternative optical source was a more likely counterpart to the X-ray source. All of these alternative optical counterparts are closer to the *Chandra* position than the counterpart chosen by Zheng et al. (2004), and most are optically fainter. Where possible (four cases) the COMBO-17 redshift estimates are used for the preferred counterpart. Otherwise the X-ray source is considered to be optically unidentified. I show the X-ray and optical images for these sources in figure 4.4. For another six of the Zheng et al. (2004) sources, the correct spectroscopic redshift is listed, but an incorrect optical position is given. These six cases were all *Chandra* sources with multiple optical counterparts listed in Szokoly et al. (2004).

The R-band magnitudes of the optical counterparts are taken from the COMBO-17 catalogue. However, in a small number of cases, this is not possible. For example, where the X-ray source lies close to a bright star. In these cases, I adopt the R-band magnitude of the optical counterpart given in Giacconi et al. (2002). I have double checked that these sources are the same as those selected by Zheng et al. (2004), and are the correct counterpart to the X-ray source.

A small number of the XMM-CDFS detections matched to 1Ms *Chandra* sources have faint ($R > 25$) optical counterparts for which photometric redshifts have been calculated by Mainieri et al. (2005). However, in each case, the Mainieri et al. (2005) redshift estimate is in agreement with the Zheng et al. (2004) value within the errors. So for simplicity and consistency, I have preferred to adopt the Zheng et al. (2004)

[1] http://www.astro.psu.edu/users/niel/hdf/hdf-chandra.html

[2] http://www.astro.psu.edu/users/niel/ecdfs/ecdfs-chandra.html

[3] http://www.mpia-hd.mpg.de/COMBO/combo_CDFSpublic.html

[4] ftp://archive.stsci.edu/pub/hlsp/gems/

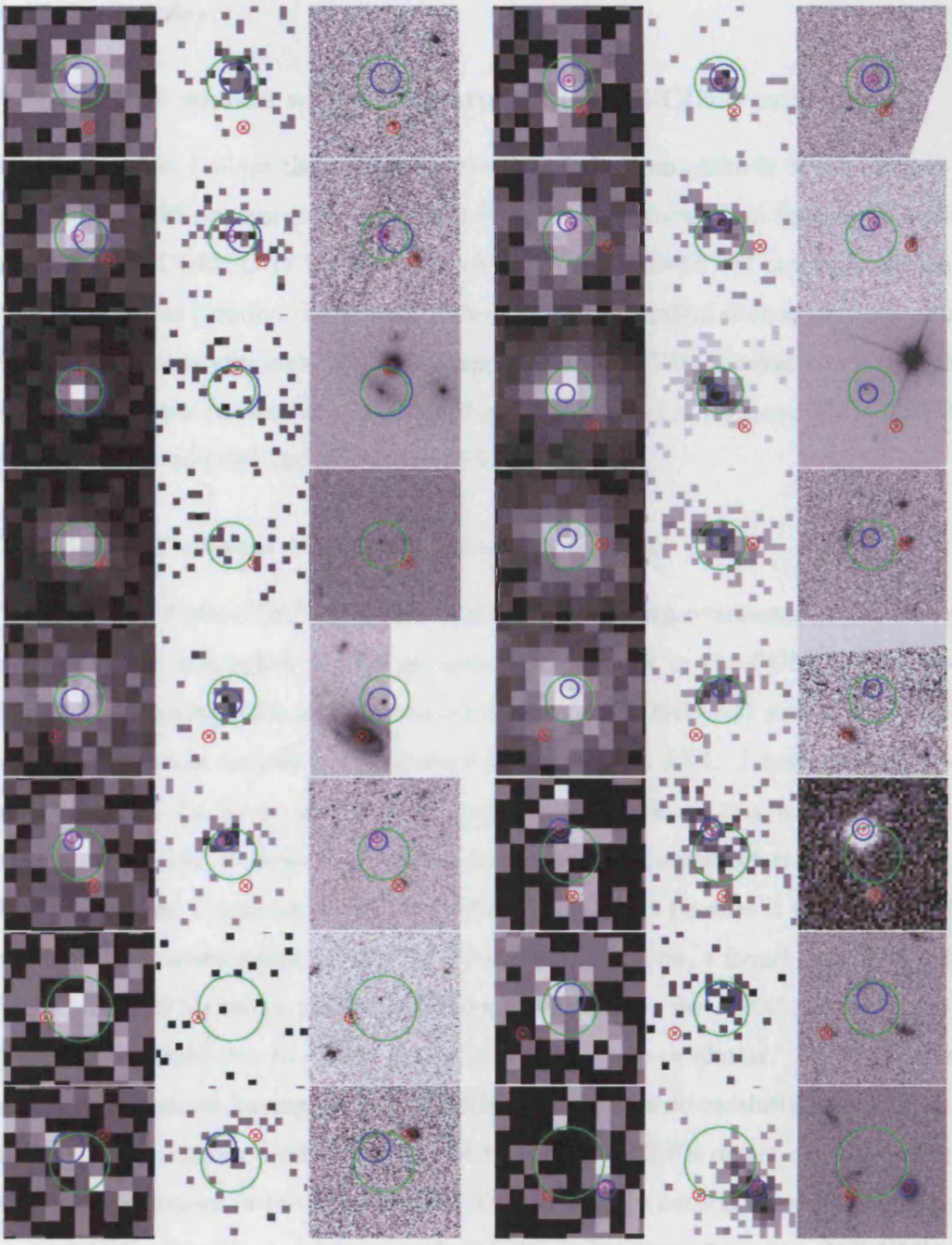

Figure 4.4: Sources with incorrect optical counterparts in Zheng et al. (2004). For each, I show the 1 Ms (left) and E-CDFS (centre) full band *Chandra* images, and the GEMS/GOODS ACS z-band image (right). The green circle marks the 1Ms catalogue position and error, the blue circle marks the E-CDFS position and error, the circled X shows the optical position given in Zheng et al. (2004). The magenta cross shows my preferred COMBO-17 optical counterpart (if any). From the top left, the images are for sources with IDs: 3, 17, 23, 25, 36, 61, 64, 70, 97, 99, 213, 517, 528, 548, 591, and 641 from Giacconi et al. (2002). For source 517 the right panel is the COMBO-17 R-band image.

redshift estimate.

XMM-CDFS sources with a counterpart in the E-CDFS catalogue

Where possible, I adopt the optical position and R-band magnitude of the counterparts to E-CDFS sources given by Lehmer et al. (2005). I then adopt the photometric redshift from COMBO-17 if there is an object in the COMBO-17 catalogue within 1″ of the optical position. In several of these cases, the optical counterpart has also been spectroscopically identified (i.e. it appears in the VVDS catalogue of Le Fevre et al. 2004, or the Szokoly et al. 2004 field galaxy list), and so for these XMM-CDFS sources I have adopted the spectroscopic identifications.

XMM-CDFS sources with no *Chandra* counterpart

There are 33 XMM-CDFS detections having no *Chandra* counterpart. For these objects I have attempted to find an optical counterpart in the COMBO-17 catalogue. The starting point was to choose the brightest COMBO-17 source inside the off axis dependent matching ellipse described in section 4.3.1. I have then manually examined the X-ray and R-band images to determine if this was the correct choice. For many of these sources, the initial choice of counterpart was adopted. For two sources, I chose an alternative optical counterpart because it was co-located with a slight enhancement in the *Chandra* image. However, I found that for 13 of the XMM-CDFS sources without *Chandra* counterparts, the *XMM-Newton* detection is most likely due to diffuse emission from a group or cluster. That is, there are several galaxies having similar COMBO-17 photometric redshifts located close to the *XMM-Newton* position. Many of these XMM-CDFS detections have their strongest detections in the softest *XMM-Newton* band (0.2-0.5 keV), supporting the group/cluster hypothesis. I have excluded these detections from the final XMM-CDFS sample, as they are unlikely to be AGN. One of the XMM-CDFS detections is located away from the centre of a low redshift, bright ($R \sim 17$) face-on spiral galaxy. I have removed this detection from the XMM-CDFS sample as it is likely to

be due to diffuse emission rather than an AGN. Finally, there are two low detection likelihood XMM-CDFS sources located very near to the edge of the *XMM-Newton* FOV, outside the E-CDFS and COMBO-17 coverage. The effective exposure time is low and the PSF is large in these regions and so it is difficult to determine if these detections are real astronomical sources. Therefore, for simplicity, I have also removed these two detections from the sample.

Figure 4.5 shows the positional offsets between X-ray positions and optical positions of sources in the XMM-CDFS sample as a function of X-ray flux.

4.3.4 A summary of the XMM-CDFS sample

After applying the *XMM-Newton* positional accuracy criteria, the detection likelihood threshold, and after removing the detections which are confused, unlikely to be point sources, or at the very edge of the field, the XMM-CDFS sample contains 309 sources. Of these, 291 have *Chandra* counterparts, and 278 are matched to an optical counterpart. A large fraction of the XMM-CDFS sources (259/309 = 84%) have optical spectroscopic identifications and/or photo-z estimates. Fifteen of the XMM-CDFS sources are associated with Galactic stars, and so are not considered further. The distribution of 0.2–10 keV X-ray flux *vs* R-band magnitude of the XMM-CDFS sources is shown in figure 4.6.

4.4 Models of the AGN N_H distribution

In the course of this chapter I test the predictions of a number of model N_H distributions, which are described below. I introduced most of these models in chapter 3, but I repeat the descriptions here for completeness. I have discarded the $R = 0$ model because it obviously does not represent the AGN population. In order to reduce the numbers of tested models, I also discard the *GSH01-R8* model which is rather similar to the "*T04*" model, which I introduce below.

The "$(\log N_H)^{\beta}$" N_H distribution models: In these models the number of AGN

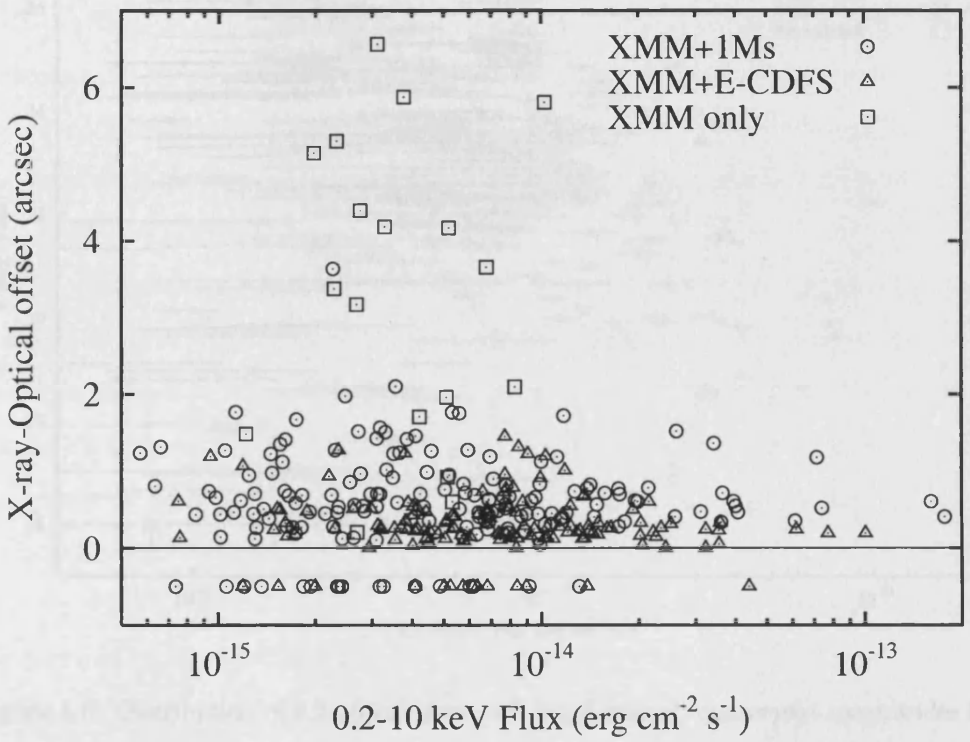

Figure 4.5: The differences between the X-ray determined position and the position of the optical counterpart, as a function of *XMM-Newton* 0.2–10 keV flux. The *XMM-Newton* detections which have counterparts in the 1 Ms *Chandra* catalogue are shown with circles, those with an E-CDFS counterpart are shown with triangles, and the XMM-CDFS sources which do not have *Chandra* counterparts are shown with boxes. Sources having no optical counterpart within the matching region are placed at offset =-0.5.

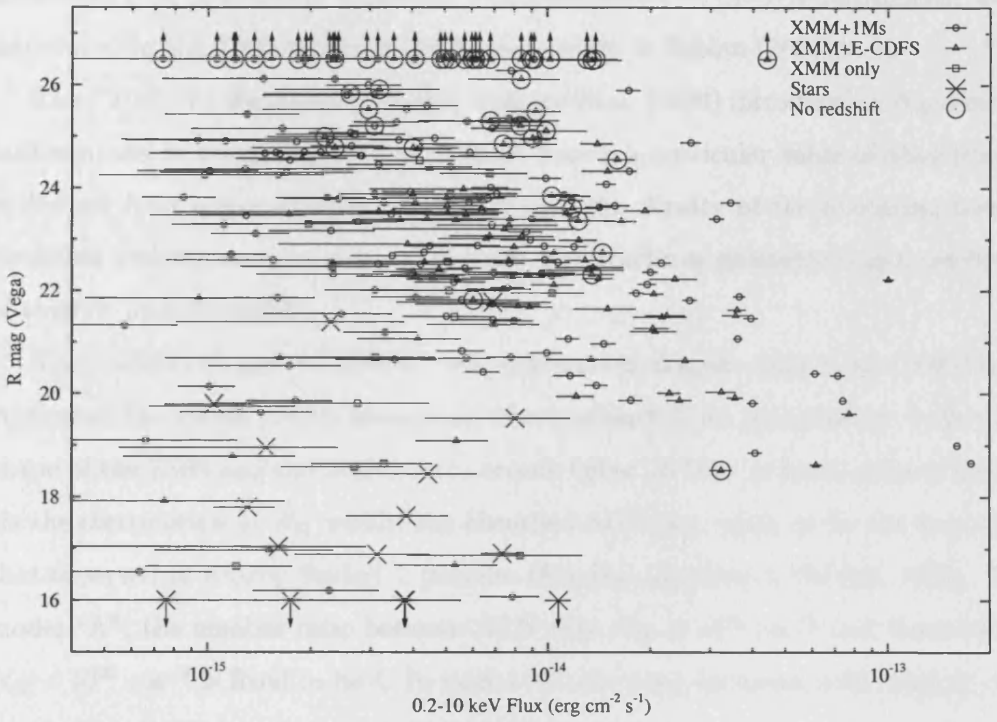

Figure 4.6: Distribution of 0.2–10 keV flux *vs* R band optical counterpart magnitudes for the XMM-CDFS sample. Objects identified as Galactic stars are indicated with crosses. XMM-CDFS detections with *Chandra* counterparts are shown with either small circles (for 1Ms matches), or small triangles (for E-CDFS matches). XMM-CDFS detections without *Chandra* counterparts are shown with small boxes. XMM-CDFS detections with no optical counterpart, or a counterpart fainter than $R = 26.5$, are marked with upward pointing arrows. Very optically bright counterparts are placed at $R = 16$, and marked with downward facing arrows. Objects with no redshift estimate are highlighted with large circles. Optical magnitudes are taken from Wolf et al. (2004), Giacconi et al. (2002), and Lehmer et al. (2005).

with absorption N_H, per unit $\log N_H$, is proportional to $(\log N_H)^\beta$, and is not dependent on redshift or luminosity. I have tested three variations by setting the parameter β to 2, 5, and 8. A similar parameterisation of the N_H distribution was introduced in the XRB synthesis models of Gandhi & Fabian (2003).

The "*T04*" N_H distribution model: Treister et al. (2004) introduce an N_H distribution model in which the fraction of AGN having a particular value of absorption is derived from a model torus. In this model, the density of the obscuring torus decreases with distance away from its plane, and the torus geometry is independent of redshift and luminosity.

The "*GSH01A*" and "*GSH01B*" N_H distribution models: Gilli et al. (2001) investigated the ability of two absorption distribution models to reproduce both the shape of the XRB and the AGN source counts below 10 keV. In both of these models the distribution of N_H within the absorbed AGN was taken to be the same as that observed in nearby Seyfert 2 galaxies (Risaliti, Maiolino & Salvati, 1999). In model "A", the number ratio between AGN with $N_H \geq 10^{22}$ cm^{-2} and those with $N_H < 10^{22}$ cm^{-2} is fixed to be 4. In model "B" the ratio increases with redshift; at $z = 0$ the ratio is 4, and at $z \geq 1.32$ the ratio is 10.

The "*U03*" N_H distribution model: Ueda et al. (2003) fitted a luminosity dependent model to the distribution of N_H in their AGN sample. In this model, the fraction of AGN having $N_H > 10^{22}$ cm^{-2} decreases linearly with luminosity, from ~ 0.6 of AGN with $L_{2-10} \leq 10^{43.5}$ erg s^{-1}, to ~ 0.4 of AGN with $L_{2-10} = 10^{45}$ erg s^{-1}.

4.5 Estimating the intrinsic properties of the sample using X-ray colours and Monte Carlo simulations

X-ray hardness ratios can be used as a surrogate for a full spectral analysis for faint X-ray sources with low numbers of counts. However, using broad band X-ray

colours means that all the information in the X-ray spectrum of a source is reduced to just a handful of hardness ratio measurements. Therefore applying this approach to observations of individual faint AGN relies on the observed sources belonging to some simple family of spectral types, for which the spectral parameters can be deduced. Spectral analyses of relatively bright AGN have shown that nearly all can be broadly described by a spectral model consisting of a primary power-law with slope $\Gamma \sim 1.9$, attenuated with some absorbing column of neutral material (Page et al., 2003; Piconcelli et al., 2003; Page et al., 2006). For such a spectral model, the absorbing column and the redshift have the greatest impact on the observed X-ray colours. There is a degeneracy between these two parameters. A heavily absorbed AGN at high redshift will have very similar X-ray colours to a low redshift, weakly absorbed AGN. If the redshifts of the AGN are known, then the degeneracy is broken; it becomes possible to use a hardness ratio based method to derive the absorption for individual AGN.

In chapter 3, I found that a better match to the X-ray colours of AGN detected in the 13^{H} deep *XMM-Newton* field was made by an AGN spectral model which included an unabsorbed cold reflection component. The net effect of the reflection component is to harden the spectrum at high X-ray energies. A number of additional spectral model components are sometimes required to provide the best fits to high signal to noise AGN spectra. However, these extra components are generally much less important than the primary power-law component when considering broad band X-ray colours. Therefore, in this chapter, I take the "absorbed power law plus reflection component" form as my baseline AGN spectral model.

The traditional approach to estimate the absorption in faint X-ray detected AGN is to fit the observed X-ray hardness ratio(s) to a model spectrum using a spectral fitting package such as XSPEC. However, there is a large degree of degeneracy in such an approach because of the number of fitted spectral parameters (N_H and/or Γ), compared to the limited number of data points. Indeed, it is typical for authors to use just a single hardness ratio measure between the 0.5–2 and 2–10 keV energy

bands. Therefore, I have devised a novel approach, in which the intrinsic properties of sources in the XMM-CDFS sample are deduced by comparing them to the "output" properties of a library of simulated AGN. This method allows account to be taken of the scatter of sources in multiple HR space (which may be strongly asymmetric, and is dependent on the observations), and so allows a rigorous estimation of confidence intervals. My approach allows all the identified sources in the XMM-CDFS sample to be treated in a consistent, uniform fashion, as opposed to examining the brighter sources using a different method to the fainter sources.

What is more, the Monte Carlo simulation process can be used to compare directly the predictions of various AGN population models against the properties of sources in the XMM-CDFS sample. Because the redshifts are known for most of the XMM-CDFS sources, I move from just a comparison of the X-ray colours of the sample (as in chapter 3), to a full comparison of the "higher level" parameters of the AGN i.e. luminosity, absorption and redshift.

I will first summarise the method used to generate a simulated library of sources, and then describe the absorption and intrinsic luminosity estimation processes.

4.5.1 Constructing the library of simulated sources

I have used the Monte Carlo method described in 2.2 to generate a large reference library of simulated sources, taking account of the exposure times and background levels in the XMM-CDFS dataset. This is very similar to the method used to simulate model AGN populations described in chapter 3. However, I decided to adopt a different baseline XLF model to that used before, namely the luminosity dependent density evolution XLF model of Ueda et al. (2003). The Ueda et al. (2003) XLF model is derived from a sample of 247 AGN detected above 2 keV taken from a variety of X-ray surveys. The XLF is calculated for the intrinsic rest frame 2–10 keV luminosity of the AGN, it is similar in form to the LDDE1 XLF model of Miyaji et al. (2000), but includes a luminosity dependence of the cutoff redshift.

Table 4.1: A summary of the Ueda et al. (2003) XLF model parameters used in this study. The normalisation parameter, A, is adjusted later to provide a better match to the source counts in the XMM-CDFS. Where appropriate, the dependence of the parameters on the Hubble constant is indicated (where $h_{70} = H_0/70$ km s^{-1} Mpc^{-1}).

Parameter name	Value	Unit
A	5.04×10^{-6}	$h_{70}^3\,\mathrm{Mpc}^{-3}$
L_*	$10^{43.94}$	$h_{70}^{-2}\,\mathrm{ergs}^{-1}$
γ_1	0.86	
γ_2	2.23	
z^*_{cut}	1.9	
L_a	$10^{44.6}$	h_{70}^{-2} ,ergs^{-1}
p_1	4.23	
p_2	-1.5	
α	0.335	

The evolution term in equation 3.7 is replaced by,

$$e(z, L_{2-10}) = \begin{cases} (1+z)^p & z < z_{cut}(L_{2-10}) \\ (1+z_{cut}(L_{2-10}))^{p_1} \left(\frac{1+z}{1+z_{cut}(L_{2-10})}\right)^{p_2} & z \geq z_{cut}(L_{2-10}) \end{cases} \tag{4.1}$$

where the luminosity dependence of the cutoff redshift $z_{cut}(L_{2-10})$ is given by,

$$z_{cut}(L_{2-10}) = \begin{cases} z^*_{cut} & L_{2-10} \geq L_a \\ z^*_{cut} \left(\frac{L_{2-10}}{L_a}\right)^{\alpha} & L_{2-10} < L_a \end{cases} \tag{4.2}$$

The best fitting parameters of the Ueda et al. (2003) XLF model are given in table 4.1, and the shape of the XLF is shown in figure 4.7. I show later that the shape of the Ueda et al. (2003) XLF model provides a reasonable match to the redshift and luminosity distributions of the AGN in the XMM-CDFS sample.

Ueda et al. (2003) also present a luminosity dependent N_H distribution model (described in 3.4.1). However, for the purposes of generating the simulated library of sources, I have chosen to use the most favoured N_H distribution from chapter 3, that is, a distribution in which the number of AGN per unit $\log N_H$ is proportional to

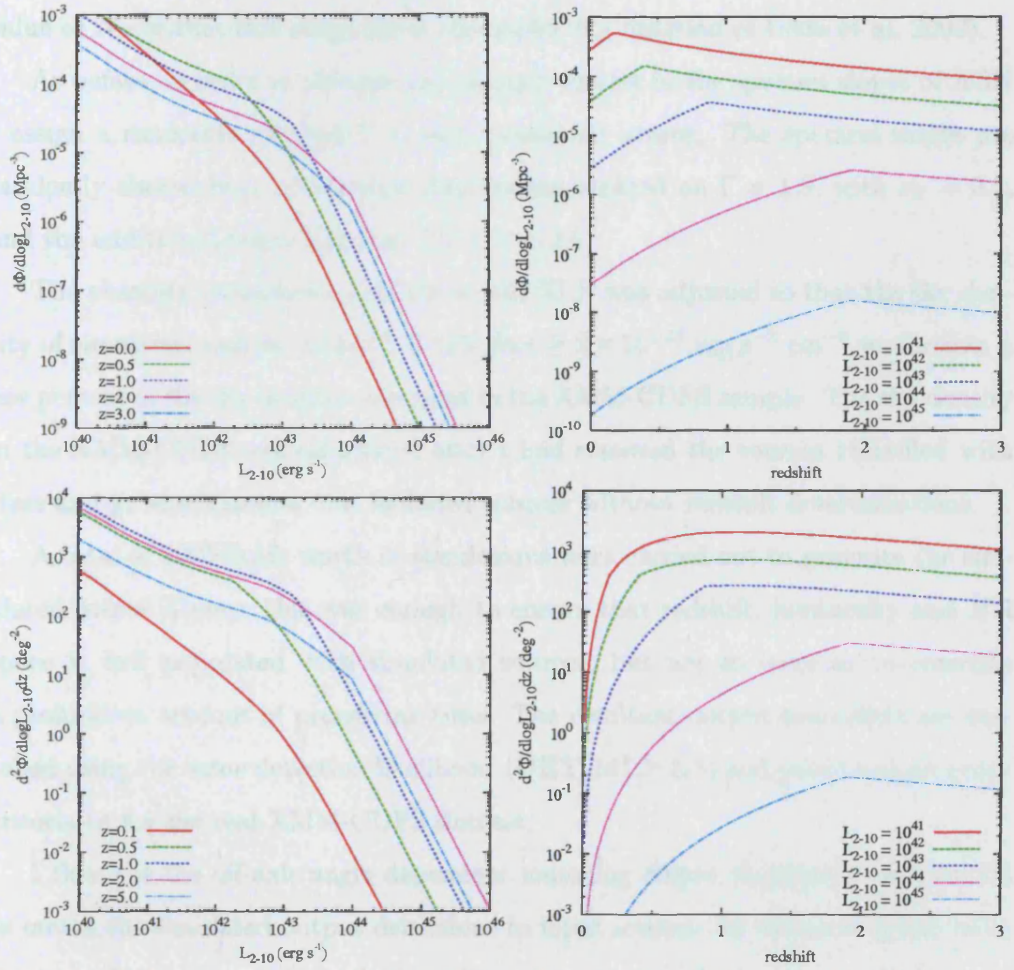

Figure 4.7: The Ueda et al. (2003) model X-ray luminosity function. The top panels show the luminosity and redshift dependence of the volume density of AGN per unit $\log L_{2-10}$. The lower panels show the luminosity and redshift dependence of the sky density of AGN per unit $\log L_{2-10}$, per unit redshift.

$(\log N_H)^8$, and is independent of luminosity and redshift. This choice of N_H model means that the simulated library sources have a broad range of absorbing columns ($10^{19} \leq N_H \leq 10^{25}$ cm^{-2}), and that the distribution is not disjoint at any particular value of N_H within this range (as is the model N_H function of Ueda et al. 2003).

As before, in order to recreate the intrinsic scatter in the spectral slopes of AGN I assign a randomly selected Γ to each simulated source. The spectral slopes are randomly chosen from a Gaussian distribution centred on $\Gamma = 1.9$, with $\sigma_\Gamma = 0.2$, and the additional constraint that $1.2 < \Gamma < 2.6$.

The absolute normalisation of the model XLF was adjusted so that the sky density of simulated sources with 0.5–2 keV flux $\geq 2 \times 10^{-15}$ erg s^{-1} cm^{-2} was within a few percent of the sky density measured in the XMM-CDFS sample. The sky density in the XMM-CDFS was calculated after I had removed the sources identified with stars and groups/clusters, but included sources without redshift determinations.

A total of 2000 fields worth of simulations were carried out to generate the simulated source library; this was enough to ensure that redshift, luminosity and HR space is well populated with simulated sources, but not so large as to consume a prohibitive amount of processing time. The resultant output sourcelists are curtailed using the same detection likelihood (DET_ML$\geq$ 8.5) and positional accuracy criteria as for the real XMM-CDFS dataset.

I then use the off-axis angle dependent matching ellipse described in section 4.3 to match the simulated output detections to input sources. In the cases where more than one input source is matched to the output source, the brightest is chosen. In order to avoid the effects of source confusion in the simulated library, the confused sources in the simulated library are identified in a similar way to the XMM-CDFS sample, i.e. any output sources which are matched to two or more input sources having comparable (within a factor of five) full-band count-rates are considered to be confused.

After applying the detection likelihood threshold, the positional criteria, and

after removing sources determined to be confused, the simulated source library contains over 500000 objects.

4.5.2 Absorption estimation technique

The measured properties of the XMM-CDFS sources which I use to estimate the absorption are the redshift (z), the vignetting corrected count rate in the 0.2–10 keV band ($R_{0.2-10}$), and three hardness ratios ($HR1$, $HR2$, and $HR3$). The absorption and luminosity of each real source are estimated from those simulated library sources which have similar values of z', $R'_{0.2-10}$, $HR1'$, $HR2'$, and $HR3'$ (here, the primed parameters are those of the output simulated sources). The following process is carried out for each optically identified source in the XMM-CDFS sample .

Firstly, the subset of objects from the simulated library which have similar redshift ($|z-z'| \leq (1+z)/10$), a similar full band count-rate, $0.5 < R'_{0.2-10}/R_{0.2-10} < 2$, and have similar hardness ratios to the real XMM-CDFS source ($|HR1 - HR1'| \leq 0.1$, $|HR2 - HR2'| \leq 0.1$, and $|HR3 - HR3'| \leq 0.2$) are selected. The weaker constraint on $HR3$ reflects the poorer counting statistics at harder energies. The possible influence that the shape of the baseline N_H distribution used to generate the simulated library may have on the estimation process is compensated for as follows. Statistical weights are calculated by counting, for a large number of bins in N_H, the numbers of library sources which satisfy the redshift and countrate criteria. The weight of each selected object is the inverse of the number counted in the N_H bin in which it lies. A 'sliding box' algorithm is then used to estimate the absorption in the real source from the N_H values and weights of the selected objects. Specifically, the sliding box algorithm searches for the value of N_H which maximises the summed weights of the selected objects lying in the interval $[\log N_H - 0.125, \log N_H + 0.125]$. This 'most likely' N_H value is taken to be the best estimate of the absorption for the real source. The confidence interval is taken to be the range of N_H about this peak which contains 68% of the statistical weight of the selected objects.

4.5.3 Intrinsic luminosity estimation technique

The intrinsic, rest-frame 2–10 keV luminosities (L_{2-10}) of the real sources are calculated using a similar technique as that used to estimate absorption. For each XMM-CDFS source, the subset of sources from the simulated library which have similar redshifts, count rates, and hardness ratios (using the same criteria as before) are selected. The differences between the $z, R_{0.2-10}$ of the real XMM-CDFS source, and the $z', R'_{0.2-10}$ of each of the selected library sources are accounted for: the L'_{2-10} of the library sources are corrected by factors of $R_{0.2-10}/R'_{0.2-10}$ and $d_L^2(z')/d_L^2(z)$ (where d_L is the luminosity distance). The median of the corrected L'_{2-10} of the selected subset of simulated sources is taken as the estimate of the intrinsic luminosity of the real source. The confidence interval is given by the range of corrected L'_{2-10} about the median value which contains 68% of the simulated subset.

4.5.4 Fidelity of the N_H/L_X estimation technique

Because this is a novel technique, I have carried out a number of tests to check that it is effective, and that the choice population model used to generate the simulated source library does not unduly bias the results. Therefore, I have measured the efficacy of the absorption/luminosity estimation technique by firstly quantifying how well it is able to estimate the N_H/L_X values of individual sources, and then I have measured how well it can recover the N_H/L_X distribution in a test population.

Ability to recover N_H/L_X of individual sources

I constructed a test population of simulated AGN following the method described in 4.5.1. The test population was generated according to the XLF model of Ueda et al. (2003), and the $(\log N_H)^8$ N_H distribution model. The test population was curtailed using the same detection likelihood threshold, positional criteria, and confusion criteria as before. The equivalent of one hundred XMM-CDFS fields were generated. The simulated output test sources were matched to input sources so that their "true"

N_H and luminosity were known. Then, for each test source, the absorption and intrinsic luminosity were measured using the N_H/L_X estimation technique, in the same way as would be carried out for real AGN. The estimated N_H/L_X values for each test source can then be compared to the "true" input parameter values.

Figure 4.8 shows the relationship between the estimated and input N_H for test sources in a number of redshift ranges. The technique recovers the input N_H values very well for test sources having moderate to heavy absorption. However, for low absorbing columns, the scatter increases rapidly. This becomes increasingly apparent at higher redshifts, as more and more of the absorption is shifted out of the EPIC bandpass. I have calculated the level below which less than 68% of test sources have N_H estimates within 0.5 dex of the input value. This ranges from $10^{21.1}$ cm^{-2} for sources in the $0 < z < 0.5$ redshift range, to $10^{22.6}$ cm^{-2} for the $3 < z < 4$ range.

The estimation technique becomes less accurate at very high levels of absorption. This is expected; for all but the highest redshift AGN having this level of absorption, virtually all the flux has been removed below 5 keV. Thus $HR1$ and $HR2$ contain little information, meaning that there is less diagnostic power to determine the amount of absorption. What is more, at these high column densities, the effects of Compton scattering, which are not included in the spectral model used here, will become significant in the spectra of the real sources. For these reasons, the most that can really be said with confidence for sources determined to have an absorbing column of greater than 10^{24} cm^{-2}, is that they have significant absorption. However, it is expected that there will be rather few of these very heavily absorbed sources in the XMM-CDFS sample because they are faint in the energy ranges where the *XMM-Newton* observations are most sensitive.

In figure 4.9 I show the relationship between estimated and input intrinsic luminosity for the test sources. The high fidelity of the technique is evidenced by the very low scatter of points about the one-to-one relation (less than ± 0.2 dex for most of the luminosity range). The luminosity estimation method is effective for both absorbed and unabsorbed test sources.

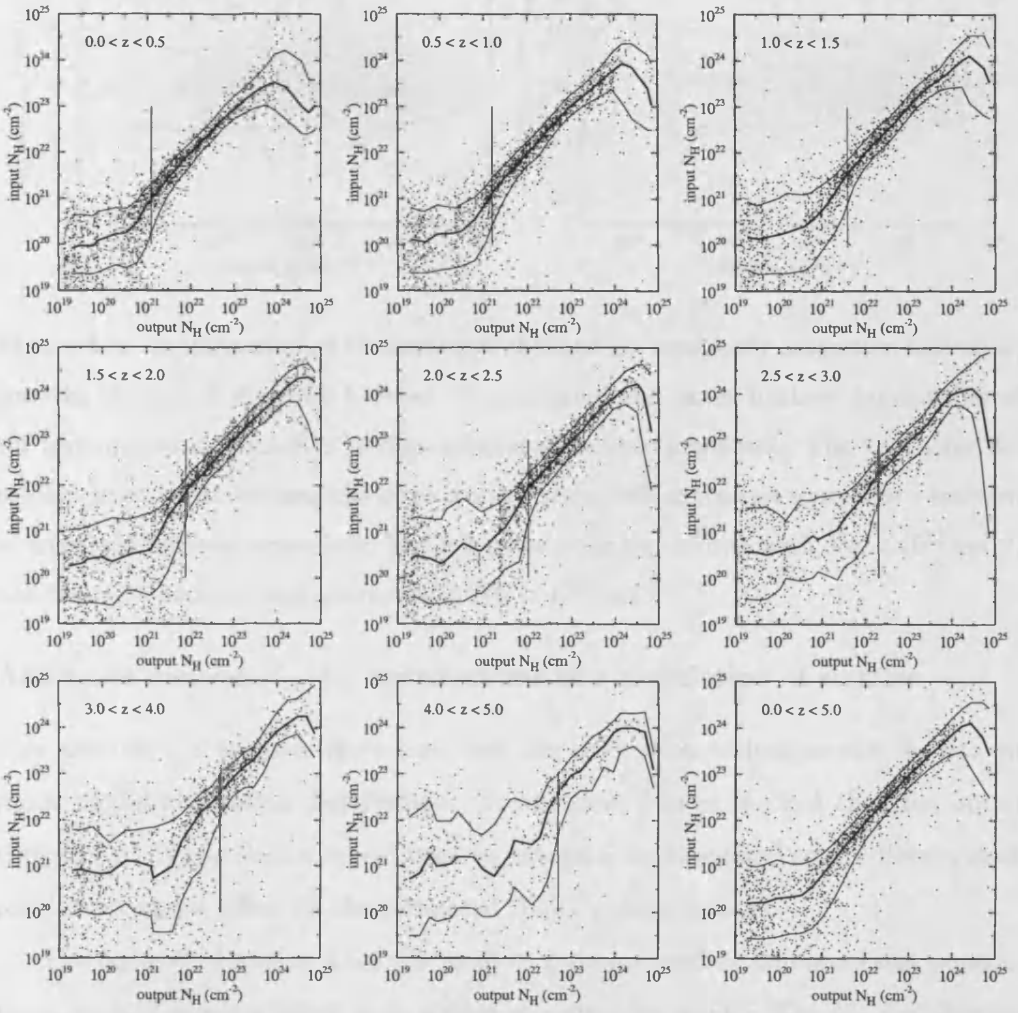

Figure 4.8: An illustration of the fidelity of the absorption estimation technique evaluated from a population of simulated test sources. The panels show the degree of scatter of output N_H about input N_H values for test sources in eight redshift bins, and for the entire redshift range. The bold curves show the median input N_H as a function of output N_H. The thin curves show the degree of scatter (they contain 68% of the test sources). The vertical lines show the lowest output N_H value for which this scatter is less than ± 0.5 dex in each redshift bin.

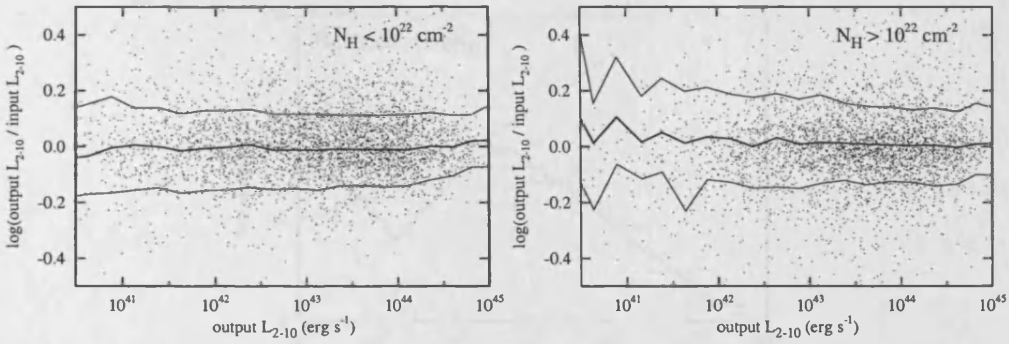

Figure 4.9: An illustration of the fidelity of the intrinsic luminosity estimation technique, showing the log of the ratio between the estimated and input intrinsic luminosities of the test sources as a function of their estimated intrinsic luminosity. The lines show the median input luminosity and the levels which contain 68% of the test sources as a function of estimated intrinsic luminosity. The left panel is for test sources with $N_H < 10^{22}$ cm^{-2}, and the right panel for test sources with $N_H > 10^{22}$ cm^{-2}.

Ability to recover N_H/L_X distributions of a population of sources

The next step is to investigate how well the estimation technique can recover an input model absorption distribution. In addition, I have checked that the initial choice of AGN population model used to generate the simulated source library does not have a major effect on the estimated N_H/L_X distributions.

The method of section 4.5.4 was used to generate several simulated test populations, each of which is based upon a different input N_H model. The N_H distribution of each population is then recovered using the N_H estimation technique. Figure 4.10 shows a comparison of the input and recovered absorption distributions for three different input N_H models. I show the input distribution of absorption in the test model, as well as the distribution in the sources that are output by the Monte Carlo simulation process. Heavily absorbed sources are less likely to be detected by the *XMM-Newton* observations than sources with lower absorbing columns; this selection effect results in the differences between the input and output distributions (see

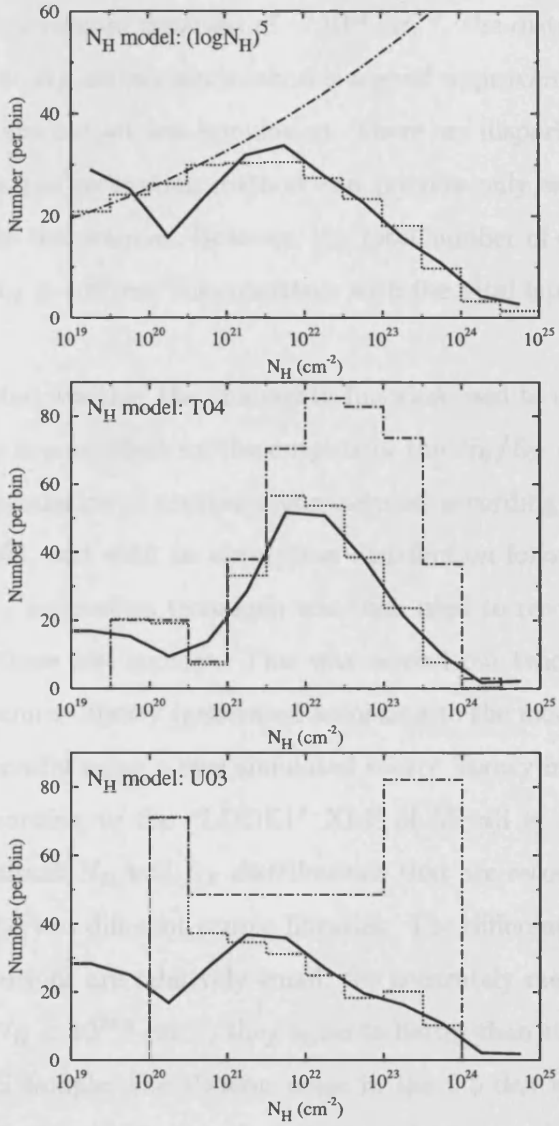

Figure 4.10: The ability of the N_H estimation process to recover the distribution of N_H in a population of test sources. The comparisons are shown for three test populations, each of which was generated using the XLF model of Ueda et al. (2003). The upper panel shows the $(\log N_H)^5$ model N_H distribution, the centre panel shows the *T04* N_H distribution, and the bottom panel shows the luminosity dependent *U03* N_H distribution. In each panel, the dot-dash line shows the shape of the input N_H model. The dotted line is the distribution of N_H in the test sources output from the Monte Carlo simulations. The solid line is the distribution of N_H which is recovered after applying the N_H estimation technique to these output test sources.

section 4.5.5). Above column densities of $\sim 10^{21}$ cm^{-2}, the distribution of absorption recovered by the N_H estimation method is a good approximation to the "true" N_H distribution in the output test population. There are disparities at low absorbing columns, where the estimation method can provide only weak constraints on the N_H values of the test sources. However, the *total* number of output test sources having estimated $N_H \leq 10^{21}$ cm^{-2} is consistent with the total number having "true" $N_H \leq 10^{21}$ cm^{-2}.

I have investigated whether the luminosity function used to construct the simulated source library has an effect on the outputs of the N_H/L_X estimation process. As before, a test population of sources was generated according to the XLF model of Ueda et al. (2003), and with an absorption distribution following the $(\log N_H)^8$ model. The N_H/L_X estimation technique was then used to recover the absorption and luminosity of these test sources. This was carried out twice, firstly using the original simulated source library (generated according to the model XLF of Ueda et al. (2003)), and secondly using a new simulated source library in which the sources are distributed according to the "LDDE1" XLF of Miyaji et al. (2000). Figure 4.11 shows the resultant N_H and L_X distributions that are recovered from the test population using the two different source libraries. The differences between the recovered N_H distributions are relatively small: for accurately measurable absorbing columns ($10^{21.5} < N_H < 10^{24.5}$ cm^{-2}) they agree to better than 10%. In comparison, in the XMM-CDFS sample, the Poisson noise in the 0.5 dex wide bins of N_H in this range is $\geq 15\%$. I conclude therefore that the N_H estimation technique is not strongly dependent on the AGN population model chosen to generate the simulated source library. However, for $L_{2-10} > 10^{44}$ erg s^{-1}, there is a noticeable difference in the luminosity distributions estimated using the two libraries (which have markedly different redshift/luminosity distributions). In order to mitigate this effect when applying the estimation technique to the real XMM-CDFS sample, it is best to use a simulated source library in which the sources have a broadly similar luminosity and redshift distribution to the sources in the XMM-CDFS sample. Therefore, for the

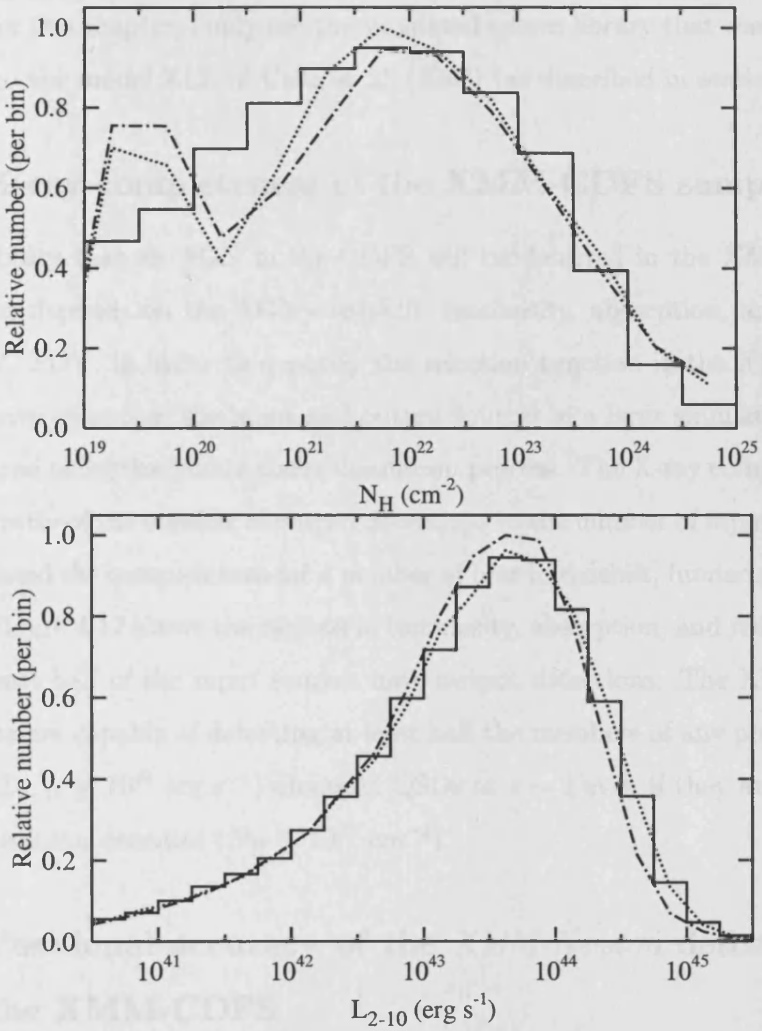

Figure 4.11: Plots demonstrating the independence of the N_H/L_X estimation technique from the X-ray luminosity function model used to generate the simulated source library. In the upper panel, the solid line shows the distribution of "true" N_H for a population of test sources, in this case generated using the $(\log N_H)^8$ model together with the model XLF of Ueda et al. (2003). The dotted line is the N_H distribution estimated when I use the simulated source library generated from the model XLF of Ueda et al. (2003). The dot-dash line shows the output distribution estimated when using a second source library which was generated from the LDDE1 XLF model of Miyaji et al. (2000). The lower panel shows the equivalent relationship for the "true" and estimated distributions of intrinsic luminosity.

remainder of this chapter, I only use the simulated source library that was generated according to the model XLF of Ueda et al. (2003) (as described in section 4.5.1).

4.5.5 X-ray completeness of the XMM-CDFS sample

The probability that an AGN in the CDFS will be detected in the *XMM-Newton* observations depends on the AGN's redshift, luminosity, absorption, and position in the EPIC FOV. In order to quantify the selection function in the XMM-CDFS sample, I have compared the input and output sources in a large simulated population generated using the Monte Carlo simulation process. The X-ray completeness is simply the ratio of the number of output detections to the number of input sources. I have calculated the completeness for a number of bins in redshift, luminosity and absorption. Figure 4.12 shows the regions in luminosity, absorption, and redshift space where at least half of the input sources have output detections. The XMM-CDFS observations are capable of detecting at least half the members of any population of luminous ($L_{2-10} \geq 10^{44}$ erg s^{-1}) obscured QSOs at $z \sim 2$ even if they are absorbed with large column densities ($N_H \sim 10^{23}$ cm^{-2}).

4.5.6 Positional accuracy of the *XMM-Newton* detections in the XMM-CDFS

I have tested the efficacy of the X-ray position matching method used to pair *XMM-Newton* detections with sources in the *Chandra* catalogues, and to optical counterparts (see section 4.3). The *XMM-Newton* imaging has been tied to the 1Ms *Chandra* sourcelist, which was in turn tied to the optical frame, and so systematic offsets have already been removed. I have examined the differences between the input and output position of the sources in the simulated library (discussed in section 4.5.1). The fraction of output detections having at least one input source within the variable matching ellipse is 98.4%. However, the input sources are simulated to a much fainter limit than that reached by the faintest output detections. Therefore

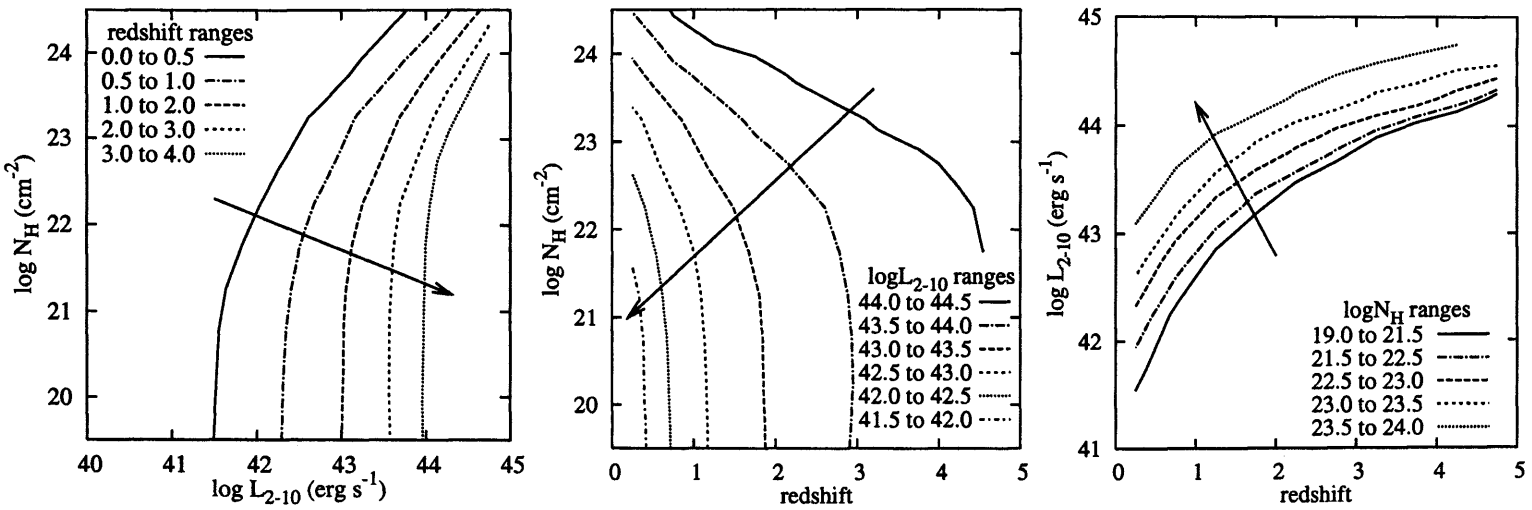

Figure 4.12: The X-ray completeness of the *XMM-Newton* observations as a function of absorption, intrinsic luminosity, and redshift. The contours are the boundaries of the regions where at least 50% of the input test population is detected in the simulated images. Left panel: The 50% completeness limits as a function of luminosity and absorption, for a number of redshift bins. Centre panel: The 50% completeness limits as a function of redshift and absorption, for a number of luminosity bins. Right panel: The 50% completeness limits as a function of redshift and luminosity, for a number of absorption bins. In each panel, the arrow serves as a guide to the eye, and indicates the direction of increasing X-ray completeness.

I added the requirement that for an input source to be considered "valid", it must have an input 0.2–10 keV countrate of at least half the output 0.2–10 keV countrate. The fraction of output detections having a "valid" input source within the variable matching ellipse is then 97.5%. Hence, for the 309 point-like detections in the XMM-CDFS sample, I expect the true astrophysical position to be within the matching ellipse for all but ~ 8 detections.

4.5.7 Direct comparison of the XMM-CDFS sample with model AGN populations

I wish to compare directly the N_H, L_X, z distribution observed in the XMM-CDFS sample with the distributions predicted by various AGN population models. To accomplish this I have used the Monte Carlo process to simulate a number of model AGN populations, then have applied the N_H/L_X estimation technique to recover the N_H/L_X distributions of the model populations. In this way the complex X-ray selection effects of the *XMM-Newton* observations can be incorporated, and the limitations of the N_H/L_X estimation technique can be accounted for. The output simulated source distributions from this process can therefore be compared like-with-like to the real XMM-CDFS sample. I have compared the XMM-CDFS sample with the predictions made by the seven different N_H model distributions described in section 4.4.

I generated simulated populations for each of the seven N_H models, according to both the XLF model of Ueda et al. (2003), as well as the "LDDE1" XLF model of Miyaji et al. (2000). For each of the fourteen combinations of N_H model and XLF model, the absolute normalisation of the XLF was adjusted in order that the simulated 0.5–2 keV source counts match the XMM-CDFS sample at 2×10^{-15} erg s^{-1} cm^{-2}. One hundred fields worth of sources were simulated for each combination of N_H model and XLF model. Finally, I have applied the N_H/L_X estimation process in order to recover the absorption and luminosity distributions

of the simulated output populations.

As discussed in section 3.4.3, some caution must be exercised when assuming that the Miyaji et al. (2000) XLF model can be scaled up to provide a good description of the whole AGN population; the same caveats apply to the use of the Miyaji et al. (2000) XLF in this chapter. Because the Ueda et al. (2003) XLF model is based on a hard X-ray selected sample, and so is less sensitive to absorption, it is expected to be more representative of the total AGN population. However, even in a 2–10 keV sample, very highly absorbed AGN ($N_H > 10^{23}$ cm^{-2}) are still selected against, and so the same caveats apply, albeit to a lesser extent. Note that rescaling the normalisation of the luminosity function for each N_H model means that the size of the integrated (over all redshifts,luminosities and absorbing columns) AGN population is different for each N_H model. Therefore it is important that any statistical tests made later compare the model populations in a fractional way, i.e. in a way that is dependent on the relative numbers of AGN with certain properties rather than the absolute number per cubic Mpc or square degree.

4.6 Results

4.6.1 Applying the N_H/L_X estimation technique to the XMM-CDFS sample

I have applied the N_H/L_X estimation technique to the members of the XMM-CDFS sample which have redshift determinations. The technique is able to evaluate the absorption and luminosity in nearly all of these sources. However, because of their very soft spectra, two AGN are not matched to any objects in the simulated source library. I discuss the properties of these sources in section 4.7.3.

4.6.2 Source counts in the XMM-CDFS

In figure 4.13 I show the differential source counts for the XMM-CDFS sample, and compare them to the predictions of several simulated model AGN populations. The shape of the predicted source count curves is dependent predominantly on the XLF model rather than on the N_H distribution. The XLF model of Miyaji et al. (2000) predicts a 0.5–2 keV source count distribution which is rather steeper than that found in the XMM-CDFS. At 0.5–2 keV fluxes above 10^{-15} erg s^{-1} cm^{-2}, the source counts predicted by the Ueda et al. (2003) XLF model are a good match to the source counts in the XMM-CDFS sample. However, the Ueda et al. (2003) XLF model underpredicts the observed numbers of sources at fainter fluxes. I carried out a Kolmogorov–Smirnov comparison of the predicted 0.5–2 keV source count distributions with the observed XMM-CDFS distribution. The KS test results demonstrate that although the Ueda et al. (2003) XLF is favoured ($P_{KS} = 0.3$), the Miyaji et al. (2000) XLF model cannot be completely rejected ($P_{KS} = 0.06$).

4.6.3 The redshift and luminosity distributions in the XMM-CDFS sample

Figure 4.14 shows the redshift distribution of the optically identified XMM-CDFS sample compared with the redshift distributions for several simulated model populations. For clarity, I have shown only the results for the $(\log N_H)^8$ and *GSH01A* models, because rather similar distributions are found in the other N_H models. It is clear that the redshift distribution predicted by the XLF of Ueda et al. (2003) is a far closer match to the redshift distribution of the XMM-CDFS sample than the prediction from the XLF model of Miyaji et al. (2000). Figure 4.14 shows that the same holds true for the luminosity distribution.

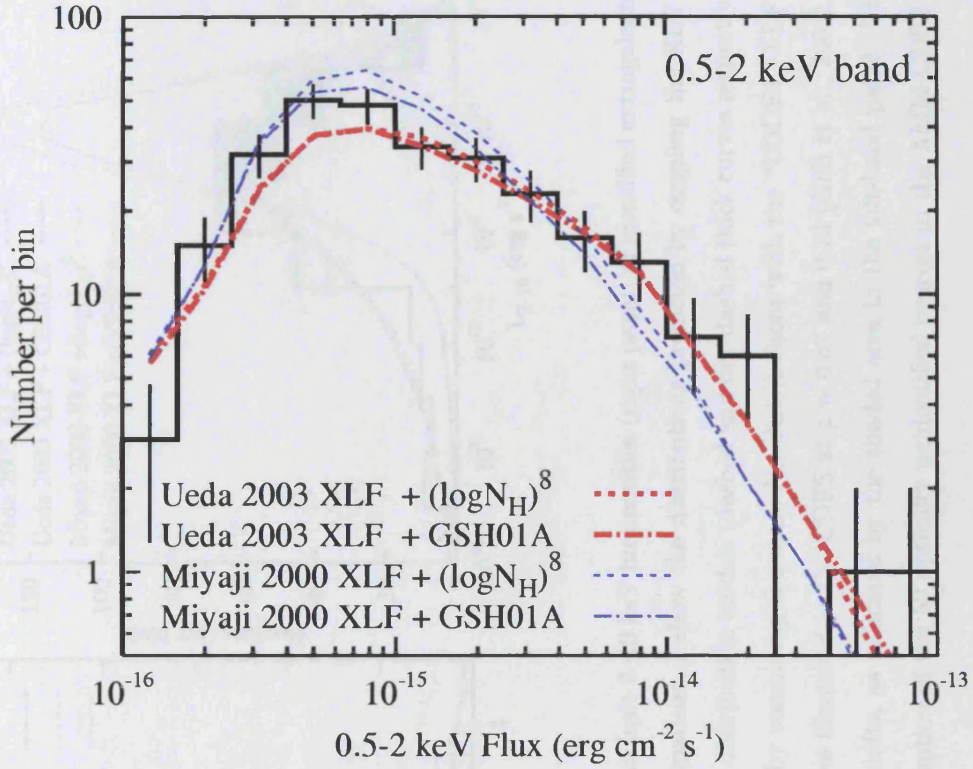

Figure 4.13: The differential source counts for the XMM-CDFS sample as a function of flux in the 0.5–2 keV energy band, shown with histograms. The source counts predicted by several different simulated model AGN populations are shown with curves. Dashed curves show the predictions when the $(\log N_H)^8$ model is combined with either the XLF model of Ueda et al. (2003, red), or with the "LDDE1" XLF model of Miyaji et al. (2000, blue). The predicted source counts from model populations generated according to the *GSH01A* N_H model are also shown (dot-dashed curves). At the faintest fluxes, the source counts are dominated by sources which have their strongest detections in other energy bands. For the purposes of this plot, confused and non-AGN sources have been removed from the sample, but unidentified sources have been included.

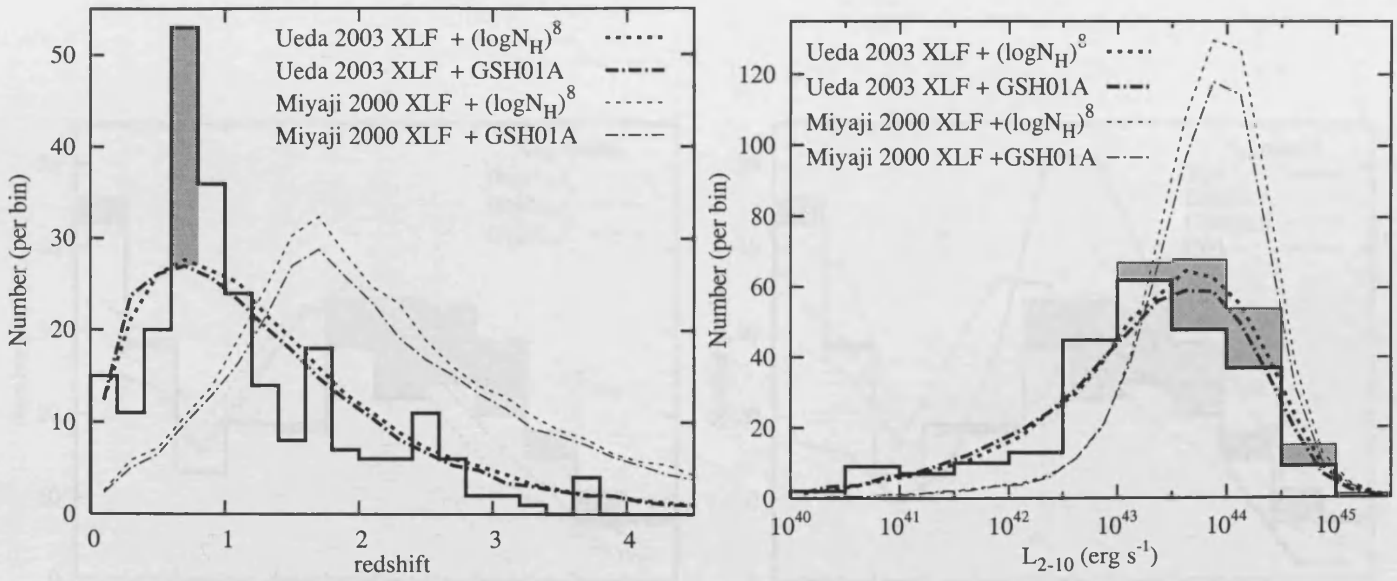

Figure 4.14: Distribution of redshifts (left panel) and intrinsic 2–10 keV luminosities (right panel) of identified extragalactic sources in the XMM-CDFS sample (solid histogram). For comparison, I show the distributions predicted by coupling the XLF model of Ueda et al. (2003) with the $(\log N_H)^8$ and *GSH01A* N_H distribution models (dashed and dot-dashed thick curves respectively). The thin curves show the rather different predictions made by combining the same N_H distributions with the "LDDE1" XLF model of Miyaji et al. (2000). There are two large-scale structures identified in the CDFS at $z = 0.67$ and 0.73 (Gilli et al., 2003). At least 26 sources in the XMM-CDFS sample lie at these redshifts, as indicated by the shaded area in the left-hand panel. The shaded area in the right hand plot shows the luminosity distribution of the 50 optically unidentified sources in the XMM-CDFS sample if they are assumed to lie at $z = 2$.

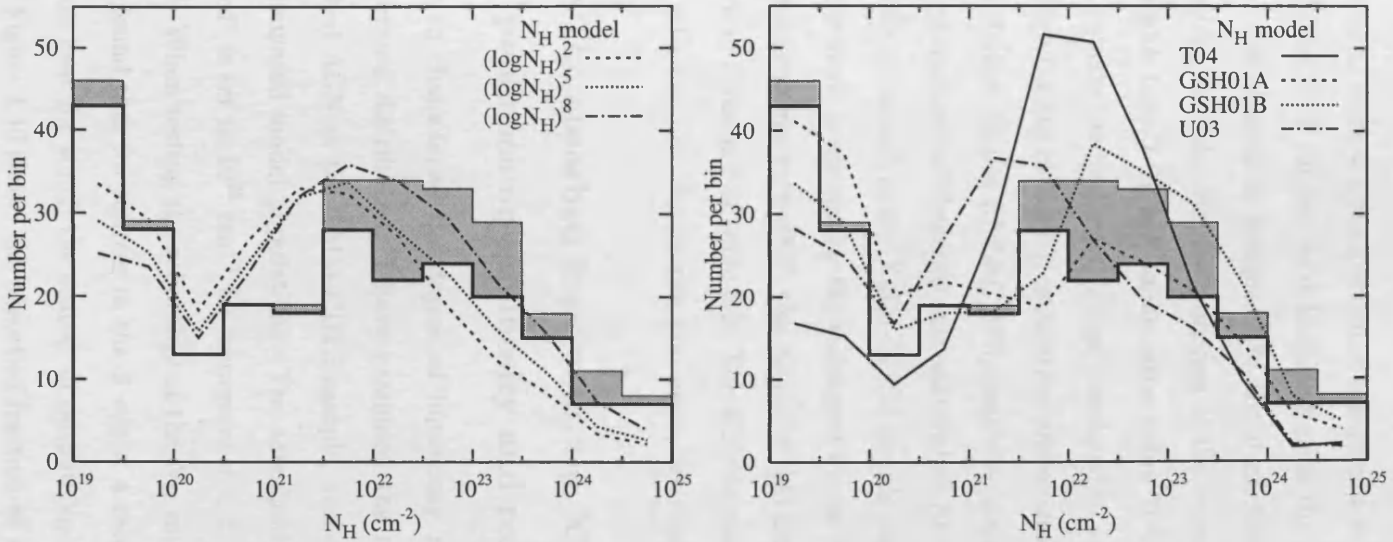

Figure 4.15: The distribution of absorption in the optically identified XMM-CDFS sample (histogram) in comparison to the predicted distributions from the seven N_H models (curves). The model N_H distributions are for simulated populations generated according to the XLF model of Ueda et al. (2003). The shaded area shows the N_H distribution of the optically unidentified sources if they are assumed to lie at $z = 2$.

4.6.4 The N_H distribution in the XMM-CDFS sample

In figure 4.15 I show the distribution of absorption in the optically identified sources in the XMM-CDFS sample, as determined using the N_H estimation technique. The measured distribution is compared to the predicted distributions from the seven simulated N_H models. For the purposes of the comparisons, the two XMM-CDFS sources which have 2–10 keV luminosities below 10^{40} erg s^{-1} are excluded, because they are outside the luminosity range simulated in the model AGN populations.

The high fidelity of the N_H estimation process means that the recovered absorption distribution in the XMM-CDFS sample contains more information than just the relative numbers of absorbed and unabsorbed AGN. The shape of the absorption distribution measured in the XMM-CDFS sample can be compared to the shape of the distributions predicted by the simulated model AGN populations. I have used the Kolmogorov–Smirnov (KS) test to make this comparison, the results of which are shown in column 1 of table 4.2. The KS test clearly discriminates between the models, with the *T04* N_H distribution being the most strongly rejected.

4.6.5 The absorbed fraction in the XMM-CDFS and its dependence on luminosity and redshift

In order to characterise the degree of luminosity and/or redshift dependence of the absorption distribution, I have examined the number ratio of absorbed and unabsorbed AGN in the XMM-CDFS sample, and compared it to the predictions of the simulated model populations. The threshold for a source to be considered "absorbed" is set to 10^{22} cm^{-2} for sources at $z < 3$, and to $10^{22.6}$ cm^{-2} for sources at $z > 3$. When testing the reliability of the N_H estimation technique (see section 4.5.4), I found that for sources in the $3 < z < 4$ range, $10^{22.6}$ cm^{-2} was the lowest absorption level for which the scatter of output N_H about input N_H was less than 0.5 dex. Figure 4.16 shows the absorbed fraction as a function of both redshift and intrinsic luminosity, in comparison to the distributions predicted from the Monte

Table 4.2: The results of statistical comparisons between the observed absorption distribution in the XMM-CDFS sample, and the predictions from a number of simulated model AGN populations. I show the results of two Kolmogorov–Smirnov (KS) comparisons between the models and the XMM-CDFS sample. Firstly, the KS test probability that the N_H distribution of the XMM-CDFS sample and the N_H distribution predicted by a model population follow the same underlying distribution. Secondly, I show the three dimensional KS test probability that the distributions of the sample and model sources in N_H, z and L_{2-10} space follow the same underlying distribution. Also shown are the results of χ^2 tests of the ability of the simulated N_H models to reproduce the redshift/luminosity dependence of the absorbed fraction in the XMM-CDFS sample. The χ^2 tests are carried out using the bins (represented by horizontal bars) in figure 4.16.

	Shape of the N_H,z,L_{2-10} distribution		Evolution of the absorbed fraction	
N_H model	N_H	N_H, z, L_{2-10}	redshift	L_{2-10}
	KS probability		χ^2/dof (Probability)	
$(\log N_H)^2$	0.002	0.0001	19 / 4 (0.0004)	18 / 4 (0.001)
$(\log N_H)^5$	0.06	0.001	9.6 / 4 (0.05)	8.5 / 4 (0.07)
$(\log N_H)^8$	0.001	0.003	6.0 / 4 (0.20)	5.5 / 4 (0.24)
T04	3×10^{-10}	<0.00001	16 / 4 (0.003)	14 / 4 (0.007)
GSH01A	0.35	0.008	6.4 / 4 (0.17)	4.8 / 4 (0.31)
GSH01B	0.001	0.003	15 / 4 (0.005)	12 / 4 (0.02)
U03	0.01	<0.00001	18 / 4 (0.001)	17 / 4 (0.002)

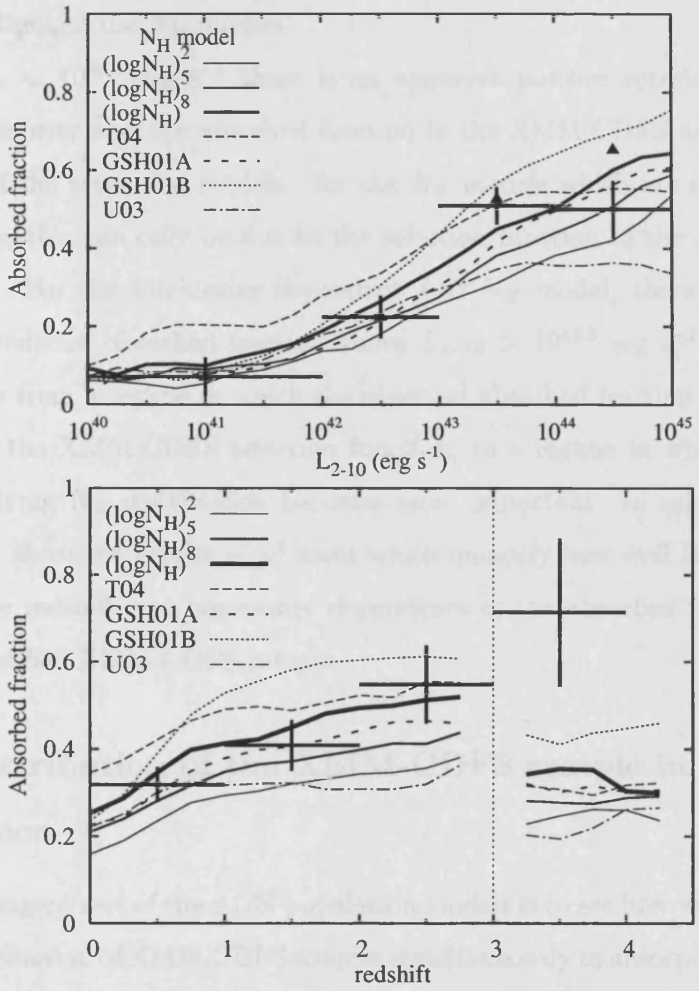

Figure 4.16: The fraction of XMM-CDFS sources with significant absorption, as a function of intrinsic 2–10 keV luminosity (upper panel), and redshift (lower panel). The threshold for a source to be considered "absorbed" is 10^{22} cm^{-2} for sources at $z < 3$, and $10^{22.6}$ cm^{-2} for sources at $z > 3$. The result for the XMM-CDFS sample is shown with errorbars: horizontal bars show the range over which the absorbed fraction has been calculated, and the vertical error-bars show the binomial uncertainty estimate, given by $\sqrt{a(1-a)/N}$, where a is the absorbed fraction, and N is the number of sources in the bin. For comparison, the curves show the absorbed fractions predicted by the N_H distribution models, calculated in bins of width 0.5 dex in luminosity (upper panel), and width 0.5 in redshift (lower panel). The triangles in the upper panel show the absorbed fraction in the XMM-CDFS if all sources without redshift determinations are placed at $z = 2$.

Carlo simulations of the N_H models.

For $L_{2-10} < 10^{43.5}$ erg s^{-1} there is an apparent *positive* correlation between intrinsic luminosity and the absorbed fraction in the XMM-CDFS sample, as well as for each of the seven N_H models. For the N_H models which are not dependent on luminosity, this can only be due to the selection function of the *XMM-Newton* observations. For the luminosity dependent *U03* N_H model, there is a levelling out of the predicted absorbed fraction above $L_{2-10} > 10^{43.5}$ erg s^{-1}. This marks the transition from a regime in which the observed absorbed fraction is determined primarily by the XMM-CDFS selection function, to a regime in which the shape of the underlying N_H distribution becomes more important. In columns 3 and 4 of table 4.2 I show the results of χ^2 tests which quantify how well the N_H models reproduce the redshift and luminosity dependence of the absorbed fraction in the optically identified XMM-CDFS sample.

4.6.6 Distribution of the XMM-CDFS sample in z,L_{2-10},N_H space

The most stringent test of the AGN population models is to see how well they reproduce the distribution of XMM-CDFS sources simultaneously in absorption,luminosity and redshift space. Figures 4.17 and 4.18 show the distribution in L_X, z and N_H of the XMM-CDFS sources and the predictions from the N_H models. By using the three-dimensional Kolmogorov–Smirnov test (3D-KS), which requires no binning, a statistical comparison can be made which interrogates the full information content of these distributions. However, the conversion from the 3D-KS statistic to a probability is rather dependent on the size of the sample and the correlations within it (Fasano & Franceschini, 1987). This is important here because of the strong correlation between z and L_{2-10} in the sample. Therefore, I have adapted the method described in chapter 3 in which the conversion from the 3D-KS statistic to a probability is calculated numerically for the actual correlation, and real number of sources

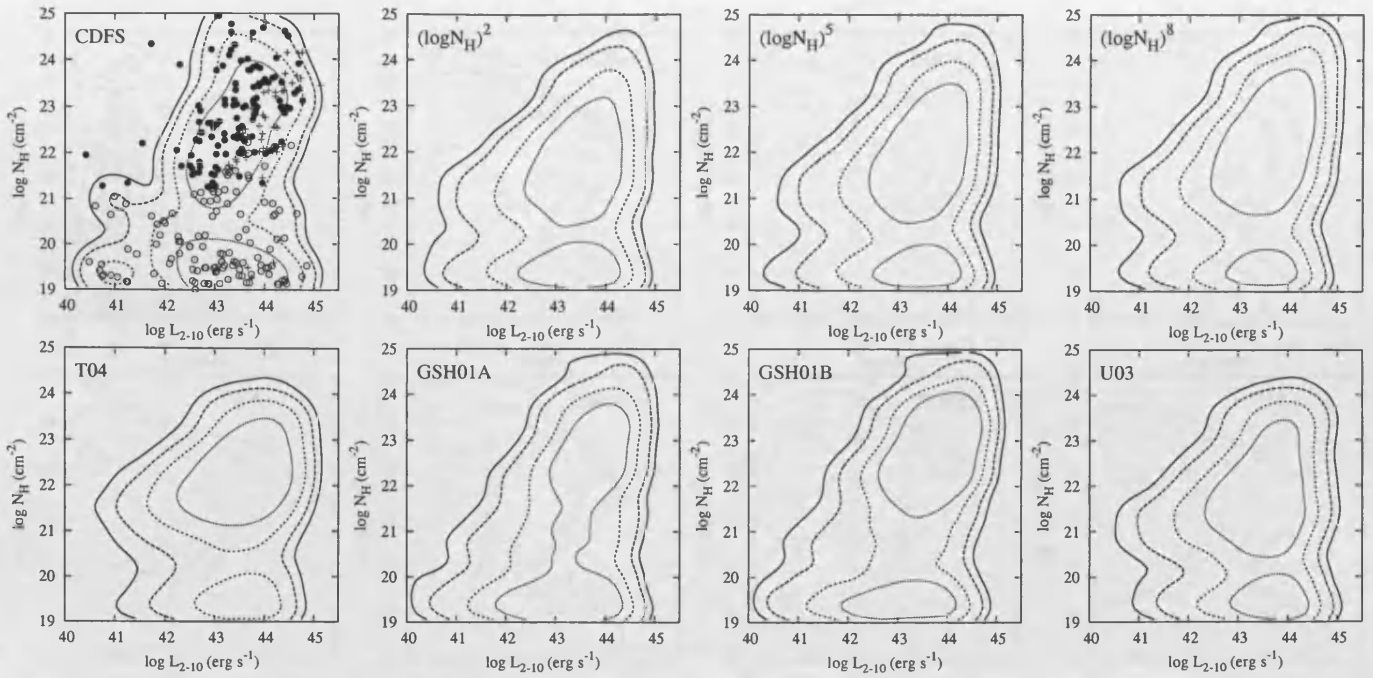

Figure 4.17: Distributions of absorption *vs* intrinsic luminosity. The distribution found in the XMM-CDFS sample is shown in the top-left panel. The other panels show the predictions from the N_H models. The contours enclose the regions occupied by 95% (solid line), 90% (long dashed line), 75% (short dashed line), and 50% (dotted line) of the sources. In order to construct the contours, the model distributions have been smoothed with a 2-D Gaussian of width 0.25 in both $\log L_{2-10}$, and $\log N_H$. A width of 0.4 is used to smooth the XMM-CDFS sample. The identified sources of the XMM-CDFS sample are also marked in the top left panel (circles). Open circles mark the sources which have absorbing columns smaller than the level where the estimation technique is accurate within 0.5 dex (see section 4.5.4). The crosses show the luminosity/absorption of the unidentified XMM-CDFS sources if they are assumed to lie at $z = 2$.

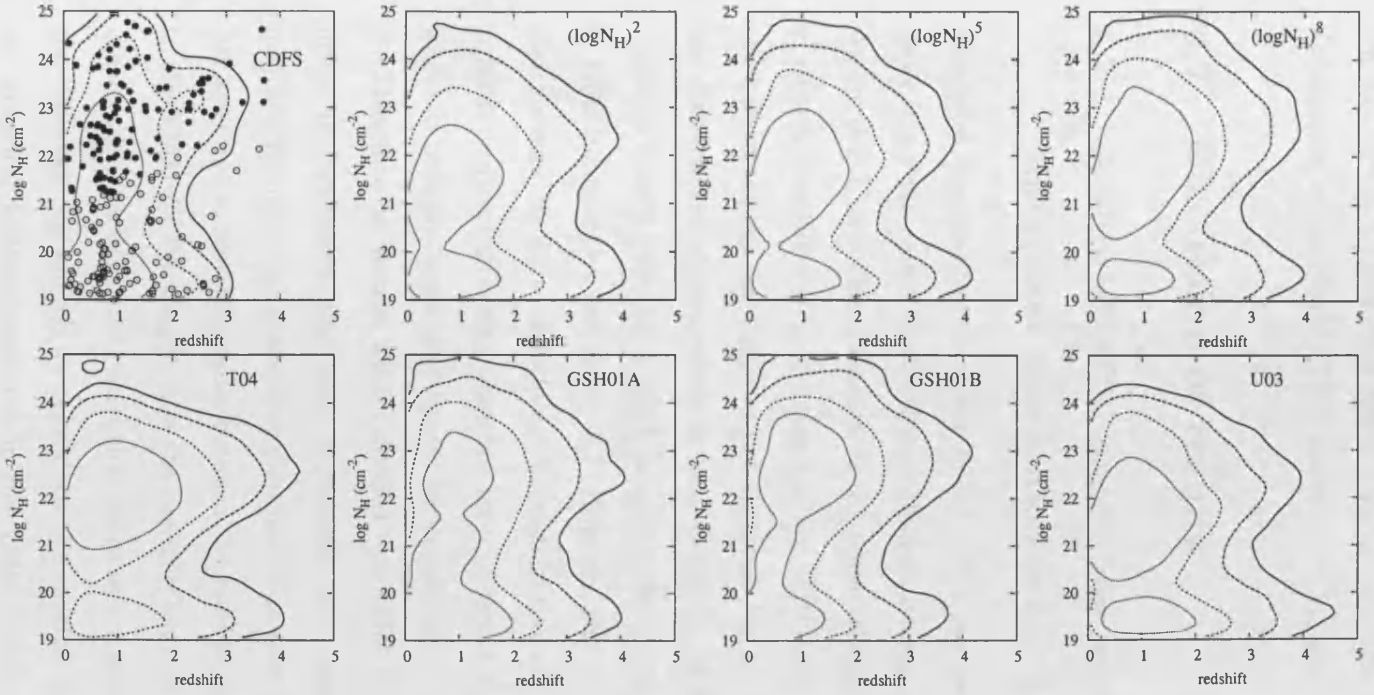

Figure 4.18: Distributions of absorption *vs* redshift. The distribution found in the XMM-CDFS sample is shown in the top-left panel. The other panels show the predictions from the N_H models. The contours enclose the regions occupied by 95% (solid line), 90% (long dashed line), 75% (short dashed line), and 50% (dotted line) of the sources. In order to construct the contours, the model distributions have been smoothed with a 2-D Gaussian of width 0.25 in both z, and $\log N_H$. A width of 0.4 is used to smooth the XMM-CDFS sample. The identified sources of the XMM-CDFS sample are also marked in the top left panel (circles). Open circles mark the sources which have absorbing columns smaller than the level where the estimation technique is accurate within 0.5 dex (see section 4.5.4).

in the tested data set. The results of the 3D-KS tests are shown in column 2 of table 4.2. It can be seen that none of the N_H models is able to reproduce fully the distribution of the XMM-CDFS sample in z, L_{2-10}, N_H space.

4.7 Discussion of results

4.7.1 Ability of AGN population models to reproduce the absorption distribution in the XMM-CDFS sample

The total N_H distribution of the XMM-CDFS sample (see figure 4.15) reveals that there is a wide range of absorbing columns present in the AGN population. The N_H models I have tested reproduce the observed distribution with varying degrees of success (see table 4.2). In table 4.2, the *GSH01A* N_H model was seen to provide the best match to the shape of the observed N_H histogram. It is remarkable that the distribution of absorption in local Seyfert-2 galaxies (on which the *GSH01A* model is based) provides a good match to the XMM-CDFS sample which reaches to QSO luminosities and to $z = 3.7$. However, this test only compares the *total* N_H distribution. So if the underlying N_H distribution of AGN is actually dependent on redshift and/or luminosity, then the total observed N_H distribution will depend on where in redshift/luminosity space the sample lies.

Therefore, in section 4.6.5 I divided the XMM-CDFS sample into several bins, firstly in luminosity, and then in redshift, and tested how well the N_H models matched the observed absorption distribution in each bin. This comparison is much less sensitive to the differences between the redshift/luminosity distributions found in the XMM-CDFS sample, and the redshift/luminosity distributions predicted by the model populations. I found that there was a marked contrast in the ability of the different N_H models to reproduce the redshift and luminosity dependence of the absorbed fraction measured in the XMM-CDFS sample. The $(\log N_H)^2$, *U03*, *GSH01B*, and *T04* N_H models are all unable to reproduce the pattern seen in the

XMM-CDFS sample. However, the $(\log N_H)^8$ and *GSH01A* N_H models provide statistically adequate (probabilities of order 0.2) fits to both the redshift and luminosity dependence of the absorbed fraction in the XMM-CDFS. The $(\log N_H)^5$ model also provides a statistically adequate match, but matches less well than the latter two models.

The $(\log N_H)^2$ and *T04* N_H models are both poor descriptions of the AGN in the XMM-CDFS sample, and are rejected with high confidence in all of the statistical tests. The former significantly underpredicts the number of absorbed AGN, and the latter significantly overpredicts the number.

The strong downturn in the absorbed fraction at high luminosities predicted by the *U03* N_H model is not seen in the XMM-CDFS sample. In addition, the *U03* N_H model predicts a total absorbed fraction of 30.1% compared to $\geq$ 38% seen in the XMM-CDFS sample. This implies that if indeed there are relatively fewer absorbed AGN at high luminosities, then the downturn can only be important at higher luminosities ($> 10^{45}$ erg s^{-1}) than are covered by this sample.

The increase in the absorbed fraction from $z = 0$ to $z = 1.3$ predicted by the *GSH01B* model is not seen in the XMM-CDFS sample. There is a suggestion that the absorbed fraction in the XMM-CDFS sample does increase at much higher redshifts ($z > 3$). However, as there are only seven XMM-CDFS sources in this redshift range, no definitive conclusion can be drawn from this dataset alone. At such high redshifts it is difficult to measure even large absorbing columns ($N_H \leq 10^{23}$ cm^{-2}) because most of the effects of the absorption are shifted out of the EPIC bandpass. Five of the XMM-CDFS objects at $z > 3$ have been spectroscopically identified by Szokoly et al. (2004): two broad-line AGN (BLAGN), and three "high excitation line" galaxies (HEX). The optical classifications of the high redshift objects tally with the X-ray determinations of their properties: the BLAGN have X-ray colours consistent with little or no absorption, but the three HEX objects are heavily absorbed. The two other XMM-CDFS objects at $z > 3$ have only photometrically determined redshifts and their X-ray colours indicate they both have significant absorption.

Of the seven N_H models examined in this chapter, the two which incorporate some redshift or luminosity dependence are both strongly rejected. The luminosity dependent *U03* N_H model predicts that there are few "type-2" QSOs, whereas the redshift dependent *GSH01B* N_H model suggests that nearly all of the accretion power at $z > 1.3$ is obscured. Neither of these scenarios fits the pattern seen in the XMM-CDFS sources, which is much better described by models in which a similar distribution of absorption is found in the AGN population at all redshifts and luminosities, namely the $(\log N_H)^8$ and *GSH01A* N_H models. Note that these two models contain similar N_H distributions to that found in the local Seyfert population (e.g. Risaliti, Maiolino & Salvati, 1999).

4.7.2 The XMM-CDFS sources without redshift determinations

There are 50 XMM-CDFS sources which do not have a spectroscopic or photometric redshift. Here I discuss the nature of these X-ray detections. For "type-1" quasars at high redshift, optical identification is made relatively easy by prominent, broad emission lines in the rest frame UV. However, for absorbed AGN, in which the optical spectrum is primarily that of the host galaxy, determination of redshifts is much more difficult. In particular, the so called "photo-z desert" ($1.5 < z < 2.5$) occurs where typical galaxy spectra have no easily identifiable features in the observed optical band. However, with the addition of near infra-red (NIR) data, this problem can be attenuated. For example, by utilising deep NIR observations, Mainieri et al. (2005) were able to photometrically identify faint ($R > 25$) counterparts to 1Ms *Chandra* sources in the CDFS sample. Indeed, Mainieri et al. (2005) showed that the optically faint objects lie on average at higher redshift than the optically brighter counterparts. The unidentified objects in the XMM-CDFS sample are nearly all optically faint; 43/50 of the unidentified sources have $R > 24.5$ (see figure 4.6). The three optically brightest of the unidentified sources do not have COMBO-17

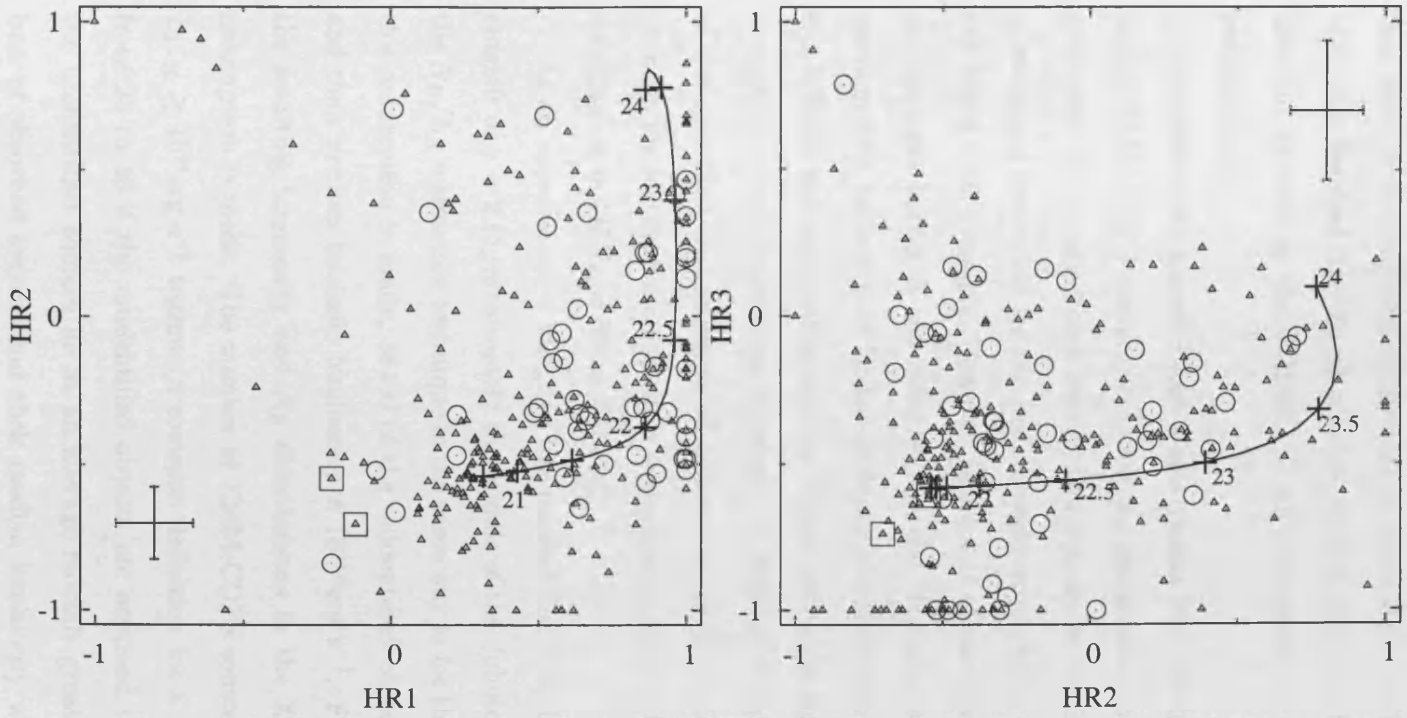

Figure 4.19: The X-ray hardness ratio distribution of the XMM-CDFS sample. The optically identified sources are shown with triangles. Sources without optical identifications are marked with circles. The two sources which do not match any objects in the simulated library are highlighted with boxes. For clarity, I show the median hardness ratio errors for the whole sample rather than the errors for each source. The solid line is the path in hardness ratio space for a model AGN lying at $z = 1$, with $\Gamma = 1.9$ and with absorbing columns ranging from zero to 10^{24} cm^{-2}. Graduations are marked, and labelled where space permits, for absorbing columns of logN_H = 24, 23.5, 23, 22.5, 22, 21.5, 21 and 19. The hardness ratios, $HR1$,$HR2$ and $HR3$ are defined in the text.

redshift estimates because they lie close to bright stars. It is reasonable to assume that most of the unidentified sources in the XMM-CDFS sample have no redshift estimates because they lie at $z > 1.5$, that is, they are beyond the upper redshift limit for galaxies in the COMBO-17 survey (where the 4000Å break has left the reddest band).

It is difficult to allow for such a bias against high redshift objects being identified in the XMM-CDFS sample, because the relationship between X-ray and optical properties of absorbed and unabsorbed sources is far from clear cut. The effect is mitigated somewhat by the high completeness (84%) of optical identification in the XMM-CDFS sample. However, figure 4.19 shows that many of the unidentified sources have harder X-ray colours than the identified sources. For example, the median $HR1$ value is 0.44 for the optically identified extragalactic sources, whereas it is 0.70 for the unidentified sources. I have tested the significance of this difference by making a two-dimensional Kolmogorov–Smirnov comparison of the distributions of the identified and unidentified sources in $(HR1, HR2)$ space. The probability that the unidentified and identified extragalactic sources follow the same underlying distribution in $(HR1, HR2)$ is 6×10^{-5}.

As an experiment, I assigned the unidentified XMM-CDFS sources a nominal redshift of $z = 2$ (approximately the middle of the "photo-z desert"), and then used the N_H/L_X estimation technique in the same way as for the identified sources. When this assumption is made, 38/50 of the unidentified objects have $N_H > 10^{22}$ cm^{-2}, and their median intrinsic luminosity is $10^{44.0}$erg s^{-1}. Figures 4.14 and 4.15 show the resulting luminosity and N_H distributions in the XMM-CDFS sample if this assumption is made. The number of XMM-CDFS sources in the $N_H \geq 10^{22}$ cm^{-2}, $L_{2-10} \geq 10^{44}$erg s^{-1} regime (a common definition for a "type-2" QSO) is doubled from 23 to 46 if the unidentified objects are assumed to lie at $z = 2$. If in fact the unidentified sources lie at an average redshift greater than 2, then the numbers of absorbed sources and their median luminosity will be higher. Figure 4.14 shows that with the addition of these luminous, high redshift AGN, the observed

redshift/luminosity distribution is closer to that predicted by the XLF of Miyaji et al. (2000). However, the unidentified sources cannot fully produce the numbers of high luminosity, high redshift AGN predicted by the XLF model of Miyaji et al. (2000).

4.7.3 XMM-CDFS sources with unusual/complex spectra

For the purposes of this study, I have considered only a rather simple AGN spectral model: an absorbed or unabsorbed power law with a component of reflected emission. However, high signal-to-noise X-ray spectra of brighter samples (e.g. Piconcelli et al., 2003; Mateos et al., 2005a,b; Page et al., 2006) have revealed that many AGN have additional spectral components. In particular, in the sample of Piconcelli et al. (2003), the spectral fits to $\sim$ 35% of the absorbed AGN were improved by the addition of an extra soft component. The most common hypotheses for the origin of this soft component is that it is either due to strong star formation activity in the host galaxy, or that it is reprocessed emission from the AGN itself. I explore the impact such additional components might have on the conclusions I have drawn from the XMM-CDFS sample.

It is possible that soft X-ray emission powered by intense star formation in the host galaxy could contribute to the spectra of some sources. In the case of an obscured AGN, an additional component will make the X-ray colours appear softer, and could lead to an underestimate of the true absorbing column. However, for X-ray luminous objects, the contribution from the star forming regions is dwarfed by the emission from the AGN. The *most* X-ray luminous starburst galaxy in the local Universe, NGC3256, has a 0.5–10 keV luminosity of $\sim 5 \times 10^{41}$ erg s^{-1} in the assumed cosmology (Lira et al., 2002). Most of the X-ray flux from NGC3256 is emitted below 2 keV. If similarly powerful starbursts are common in AGN host galaxies, then one would expect to underestimate N_H for some AGN which have 2–10 keV luminosities below $\sim 10^{42}$ erg s^{-1}. The effect will be less pronounced at

high redshifts where the peak of the thermal star forming component is shifted to lower energies. For AGN with $L_{2-10} > 10^{42}$ erg s^{-1}, any contribution to their X-ray colours from star formation will be dwarfed by the primary AGN component. In fact, nearly all (88%) of the optically identified sources in the XMM-CDFS sample have intrinsic 2–10 keV luminosities above 10^{42} erg s^{-1}, and so star formation should not be very important.

X-ray emission from AGN can reach the observer indirectly via reprocessing in a body of photo-ionised plasma, which extends well beyond the obscuring "torus" (e.g. Turner et al., 1997). For AGN where much of the direct X-ray continuum is absorbed, the reprocessed emission may constitute a large fraction of the observed soft X-ray flux. For example, in the archetypal Seyfert-2 galaxy, NGC1068, virtually all of the soft X-ray flux can be attributed to reprocessed emission from the AGN (Kinkhabwala et al., 2002). However, the intensity of the scattered component is generally only a small fraction of the primary power law component (Turner et al., 1997; Page et al., 2006). So unless the AGN is very heavily absorbed (like, NGC1068 for example), the direct power law component will still dominate the observed spectrum.

A very small number of the AGN in the *XMM-Newton* samples of Piconcelli et al. (2003) and Page et al. (2006) have X-ray spectra which are best fitted by a model in which the absorbing material is ionised. The net effect is an X-ray spectrum which has a deficit of flux at intermediate energies (at around a few keV, see section 1.2.2) compared to a simple power law. Even with high signal to noise X-ray spectra, and moderate energy resolution, it is often hard to distinguish between an ionised absorber, and an AGN with a neutrally absorbed power-law component plus an additional soft component. I do not expect to be able to differentiate between these cases in this work because I have examined only broad band X-ray colours.

When applying the N_H/L_X estimation method to the XMM-CDFS sample, there were two sources which did not match any of the objects in the simulated source library. These two objects have X-ray colours which are not well matched to the

"power law plus reflection" model spectrum, possibly because they have more complex spectra. One can estimate the number of such non-matches which are a result of the finite length of the simulated source library (i.e. not all of z,L_X,N_H,Γ space is populated with library sources). Indeed, in the simulated test populations there are on average 0.8 sources per field which remained unmatched. This rate is not far off the 2 unmatched sources found in the XMM-CDFS sample. However, examination of the unmatched simulated sources reveals that they are mostly heavily absorbed, high redshift objects detected in the simulated images at low signal to noise. In contrast, the two unmatched sources in the XMM-CDFS are among the X-ray brightest in the sample. Their spectra are rather soft compared to the other sources in the XMM-CDFS sample (see figure 4.19). The first source is identified with a broad line AGN at z=1.031 (CDFS-044, Szokoly et al. 2004). The second source is identified with an optically bright galaxy with photometric redshift 0.539 ± 0.03 (E-CDFS-381, Lehmer et al. 2005; Wolf et al. 2004). These two unmatched sources both have $HR1 \leq -0.1$, softer than the bulk of the unabsorbed objects in the sample. However, they both have relatively normal HR2 and HR3 values, indicating that their spectral slopes flatten toward higher energies. For the purposes of the statistical tests earlier, I assumed that these two objects have zero absorption and calculated their rest frame 2–10 keV luminosities from their observed 2–5 keV flux assuming a photon index of 1.9.

There may be other examples of these soft, unabsorbed AGN in the XMM-CDFS sample, which have less pronounced spectral shapes and/or are measured with lower signal to noise. However, the soft spectra of such objects will not prevent my N_H/L_X estimation method from correctly determining such objects to be unabsorbed AGN.

In summary, I expect these extra spectral components to have only a small influence on the analysis of the N_H distribution in the XMM-CDFS sample. Although additional soft components may be a common feature in absorbed AGN, their amplitudes are small in comparison to the primary power law component. The net effect will be that the absorbing columns are underestimated for some sources. For some

sources, an ionised absorber could cause the N_H estimation scheme to miscalculate the absorption, but the expected numbers of such objects are small.

4.8 Regions of extended emission in the *XMM-Newton* observations of the CDFS

There are four prominent regions of diffuse soft X-ray emission in the *XMM-Newton* observations of the CDFS, which are highlighted in figure 4.1. It is beyond the scope of my work to investigate these sources in depth, because this diffuse emission is most likely due to clusters of galaxies and not AGN. I have compared the redshift distributions of COMBO-17 galaxies in each of these four (manually determined) ellipsoidal regions to the redshift distribution over the whole field. Region #2 (located at RA $03^H32^m45.8^s$, Dec -27°40′ 57.3″) has a very clear peak in its redshift distribution at $z \sim 0.75$. Region #4 (located at RA $03^H31^m49.7^s$, Dec -27°49′ 21.7″) has several weak peaks, but with the most prominent lying at $z \sim 0.67$. For regions #3 and #1 (located at RA $03^H32^m25.8^s$, Dec -27°58′ 55.6″ and RA $03^H33^m21.2^s$ Dec -27°48′ 53.4″ respectively), the strongest peaks lie at $z \sim 0.14$. In fact this redshift bin is enhanced over the entire COMBO-17 field, suggesting that these diffuse objects are embedded in a sheetlike structure with angular extent equal to or greater than the $0.5° \times 0.5°$ COMBO-17 field. I note that extended source #1 matches CXOECDFS J033320.3-274836 of Lehmer et al. (2005), and that extended source #4 was detected in the 1Ms *Chandra* imaging (source #645 of Giacconi et al., 2002).

4.9 Comparison to the *Chandra* spectral analysis of Tozzi et al. (2006)

Very recently, Tozzi et al. (2006) have published the results of a *Chandra* X-ray spectral analysis of the sources in the 1Ms *Chandra* catalogue. In figure 4.20 I

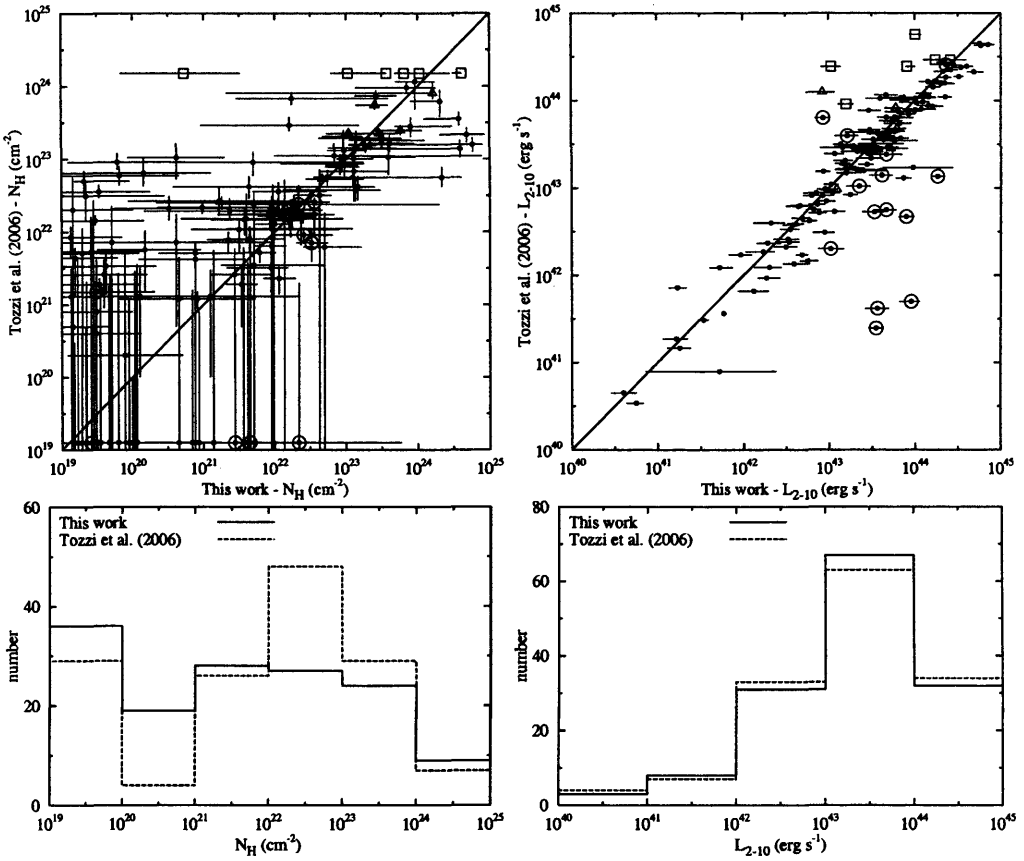

Figure 4.20: A comparison of the N_H and L_{2-10} values for the 157 AGN which appear in both the XMM-CDFS sample, and the Tozzi et al. (2006) sample. The sources determined to have zero N_H by Tozzi et al. (2006) are plotted at $N_H = 10^{19.1}$ cm^{-2}. Sources treated by Tozzi et al. (2006) as "Compton Thick", and so fitted with a pure reflection spectrum, are marked with open boxes. Sources determined to have an additional soft component by Tozzi et al. (2006) are marked with triangles. Some of the redshifts in Tozzi et al. (2006) are taken from Zheng et al. (2004). I have highlighted with circles the sources which I believe to have incorrect optical counterparts, and therefore probably wrong redshifts. Lower panels show the equivalent histograms of absorption and luminosity. The sources which I believe to have incorrect redshifts have been excluded for the histograms.

compare the N_H and L_{2-10} of the 157 AGN which appear in both the XMM-CDFS sample, and the Tozzi et al. (2006) sample. The Tozzi et al. (2006) spectral fitting has been carried out over the 0.7–7 keV range, and so is relatively insensitive to intermediate absorption ($10^{21} < N_H < 10^{22.5}$ cm^{-2}), especially at high redshift. There are a small number of sources which my estimation technique finds to have $N_H > 10^{24}$ cm^{-2}, but for which the Tozzi et al. (2006) measurement is more than a decade smaller. Assuming the Tozzi et al. (2006) estimate is correct, then this effect is similar to that seen in the simulated test populations (see section 4.5.4), and is an intrinsic limitation of my N_H estimation process. Only a small number of sources are affected, and so I do not expect this to alter the results of this chapter. Several of the sources which I determine to have effectively zero absorption ($N_H < 10^{21}$ cm^{-2}) are found to have a high ($N_H > 10^{22}$ cm^{-2}) absorbing column by Tozzi et al. (2006). However, as noted by these authors, the limited soft energy range of the *Chandra* data can cause low absorption sources to appear significantly absorbed, especially toward high redshifts.

The luminosity measurements agree very well as demonstrated by figure 4.20. The exceptions are the sources for which I have chosen an alternative (or no) optical counterpart than that chosen in Zheng et al. (2004).

4.9.1 Why have other X-ray surveys arrived at different conclusions?

Many authors have reported a lack of X-ray selected absorbed AGN at high redshifts and with high luminosities, and have developed luminosity dependent absorption schemes to explain this phenomenon (e.g. Franceschini et al. 2002; Steffen et al. 2003; Ueda et al. 2003; Barger et al. 2005; La Franca et al. 2005; Lamastra, Perola & Matt 2006). There are a number of possible explanations for why my findings are so different to these other studies.

Spectroscopic incompleteness and misidentifications

The 2–8 keV band luminosity function derived by Barger et al. (2005) relies at its faint end on an X-ray sample which is only 50–60% spectroscopically identified. Photometric redshifts raise their identified fraction, but will systematically miss objects in the 1.5–2.5 redshift interval. I have shown that the 50 XMM-CDFS sources without redshifts have harder than average colours, and therefore could be intrinsically luminous but heavily absorbed QSOs. It is possible therefore that the Barger et al. (2005) sample is underestimating the size of the population of such objects. This effect could explain the lack of objects in their sample lying at $z > 1$, and having *observed* 2–8 keV luminosities below 10^{44} erg s^{-1}.

Manual examination of the optical and *Chandra* imaging in the CDFS (in particular, the high spatial resolution, and very sensitive GEMS 'z'-band images) reveals that 16 of the optical counterparts chosen by Zheng et al. (2004) for the 1Ms *Chandra* sources are probably incorrect (see section 4.3.3). The preferred counterparts, which are generally optically faint, may well lie at higher redshifts than the host galaxy redshifts determined by Zheng et al. (2004). I have not examined all the 1 Ms *Chandra* sources, just those which are matched to *XMM-Newton* sources, so there may be additional cases of misidentification in the Zheng et al. (2004) sample.

X-ray selection function

The faintest sources in the sample of Ueda et al. (2003) are taken from the first 1Ms observations of the *Chandra* Deep Field North. A 2–8 keV flux limit of 3.0×10^{-15}erg s^{-1} cm^{-2} was applied to define the Ueda et al. (2003) sample (much shallower than the limit of the *Chandra* data). Many of the high redshift, absorbed sources in the XMM-CDFS sample would not have been selected with this criterion. To confirm this, I extrapolate the 2–5 keV band fluxes of the XMM-CDFS sources to the 2–8 keV band, assuming a conservatively hard power law slope of $\Gamma = 1.4$. Then, of the XMM-CDFS sources with $z > 1$ and $N_H > 10^{22}$ cm^{-2}, most (29/52)

have 2–8 keV fluxes below the flux limit of the Ueda et al. (2003) sample. Therefore it is unsurprising that by extrapolating the XLF and N_H model of Ueda et al. (2003) to fainter flux limits, it has not been possible to reproduce fully the XMM-CDFS sample.

Correction of luminosity for absorption effects

In X-ray studies of the absorbed AGN population, it is important that the fluxes, and hence luminosities, are calculated without correction for absorption. For high column densities, this introduces an underestimate of the intrinsic luminosity. For example, fluxes calculated from the *Chandra* 2–8 keV band are dominated by counts at the soft end of this range. For sources at $z < 3$, and for columns higher than $N_H = 10^{23}$ cm^{-2}, the effect of absorption can therefore be large. In figure 4.21, I show the difference between the intrinsic 2–10 keV rest frame luminosities estimated using my Monte-Carlo method, and the luminosities calculated directly from the observed 2–5 keV fluxes (calculated using equation 3.5, assuming a $\Gamma = 1.9$ power law). The mean luminosity correction factor for sources with $N_H \geq 10^{23}$ cm^{-2} is 2.2. Three of the "type-2" QSOs in the XMM-CDFS sample would have had apparent luminosities below 10^{44} erg s^{-1} if I had not corrected for absorption.

Sensitivity to absorption at high redshifts

XMM-Newton EPIC is sensitive to a lower photon energy limit ($\sim$ 0.2 keV) than *Chandra* ACIS-I. This is particularly useful in parts of the sky with a very low Galactic column (such as the CDFS). Section 4.5.2 describes how I took advantage of this additional bandwidth by including data from the 0.2–0.5 keV band in the N_H estimation technique. At low redshifts, the $HR1$ is sensitive to columns of $N_H \geq 10^{21.1}$ cm^{-2}. More importantly, for redshifts up to ~ 3, the estimation method can reliably detect columns of $N_H \geq 10^{22}$ cm^{-2}, which is the traditional dividing line between absorbed and unabsorbed AGN (see section 4.5.4). Surveys with *Chandra* may therefore class many absorbed AGN at high redshift as unabsorbed objects.

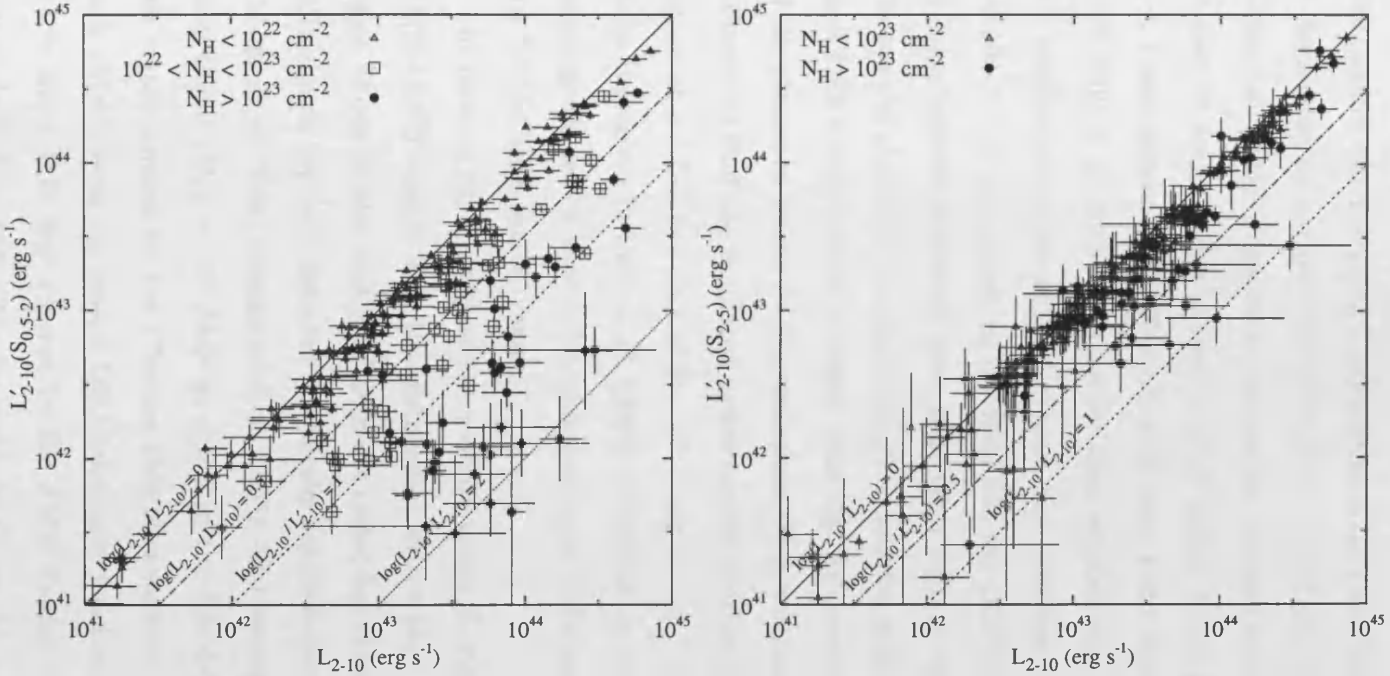

Figure 4.21: A comparison of the luminosities of XMM-CDFS sources calculated using my Monte-Carlo method, L_{2-10}, with the luminosities derived from the observed 0.5–2, and 2–5 keV fluxes with no correction for absorption, $L'_{2-10}(S_{0.5-2})$, and $L'_{2-10}(S_{2-5})$ respectively. The $L'_{2-10}(S_{0.5-2})$ and $L'_{2-10}(S_{2-5})$ values were calculated from the observed fluxes assuming a $\Gamma = 1.9$ power law. One can see that even in the 2–5 keV band, AGN with columns $N_H > 10^{23}$ cm^{-2} have a significant fraction of their flux removed by absorption.

4.9.2 Large scale clustering in the CDFS

The redshift distribution of X-ray sources in the 1Ms *Chandra* catalogue in the CDFS is dominated by narrow over-densities at z=0.67 and 0.73 containing $\geq$38 sources (Gilli et al. 2003). These over-densities are also seen in the XMM-CDFS sample with at least 26 sources lying in these redshift spikes. I note that large redshift spikes at $z \leq 1$ also appear in the 2Ms *Chandra* Deep Field North catalogue (Barger et al., 2003; Gilli et al., 2005). It is not yet clear whether such clustering of AGN at $z \leq 1$ is a ubiquitous feature of the Universe, and therefore common over the whole sky. Despite the enhancements at low redshifts, the CDFS field has a total sky density of X-ray sources somewhat lower than other deep X-ray fields (Rosati et al., 2002; Manners et al., 2003; Nandra et al., 2005; Loaring et al., 2005). This suggests that the CDFS is under dense at higher redshifts. As described in section 4.5.7, for each of the simulated model AGN populations the normalisation of the XLF model was adjusted so that the simulated source counts matched the observed integral source counts at a 0.5–2 keV flux of 2×10^{-15} erg s^{-1} cm^{-2}. An XLF normalisation 0.7 times that given in Ueda et al. (2003) is required in order for the simulated source counts predicted by their full population model (XLF and *U03* N_H model) to match the XMM-CDFS source counts.

In figure 4.22, I compare the redshift versus B magnitude distribution of the XMM-CDFS sample with the distribution of the sources in the 1Ms CDFS catalogue. It can be seen that the CDFS has rather few of the optically bright, high redshift objects typically detected in optically selected quasar surveys (e.g. 2QZ/SDSS Croom et al. 2004; Abazajian et al. 2003). For example, there is only one optically bright ($B_{AB} < 21$) AGN at $z > 2$ in the XMM-CDFS sample, and none in the region covered by the *Chandra* 1Ms observations. The predicted numbers of such objects from the optical QSO luminosity function of Croom et al. (2004) are 3.1 in the $\sim$ 0.18 deg^2 covered by the *XMM-Newton* observations, and 1.9 in the $\sim$ 0.11 deg^2 of the 1Ms *Chandra* coverage. Given this mean sky density, there is an

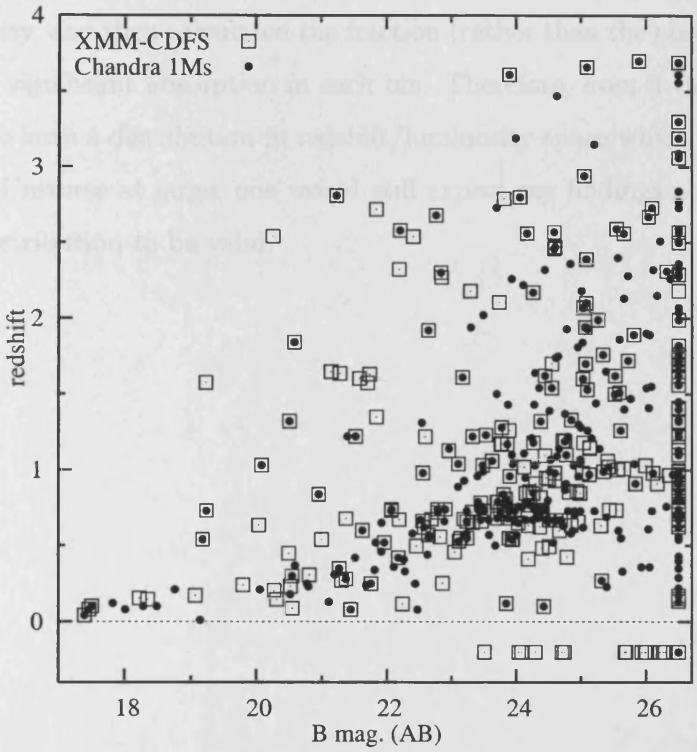

Figure 4.22: The redshifts and B band magnitudes of the extragalactic objects in the XMM-CDFS sample (squares), and the 1Ms *Chandra* CDFS catalogue (filled circles). The optical magnitudes have been calculated by matching the X-ray sources to objects in the EIS catalogue (Arnouts et al., 2001).

82% probability that more than one $B_{AB} < 21$ quasar at $z > 2$ would be observed in the XMM-CDFS. This is again consistent with the XMM-CDFS field being an under-dense region of the sky at high redshifts.

I note that in the statistical comparison of the models and the XMM-CDFS sample in section 4.6.5 I first grouped the AGN into a number of bins in redshift/luminosity, and then calculated the fraction (rather than the absolute number) of AGN with significant absorption in each bin. Therefore, even if the AGN in the CDFS field *do* have a distribution in redshift/luminosity space which is unrepresentative of the Universe at large, one would still expect my findings about the AGN absorption distribution to be valid.

Chapter 5

The *XMM-Newton* Deep One Degree Survey: A large sample of faint X-ray selected AGN

5.1 Introduction

Over the last couple of years deep *XMM-Newton* observations of a number of extragalactic fields have been made, and these data are now available in the public archive. In this chapter I extend the techniques introduced in chapter 3 to a much larger sample of X-ray selected AGN detected in *XMM-Newton* fields. I consider the *XMM-Newton* observations of the 13^{H}, Lockman Hole (LH), CDFS, CDFN, LBQS-2212 and 1^{H} fields together. The total sky area covered by these data is more than a square degree. I show later that over 1800 X-ray sources are detected in these observations, $>$ 8 times the sample size that I used in chapter 3. What is more, the observations in the LH, CDFS, CDFN, and LBQS 2212 fields are deeper than those in the 13^{H} field, meaning that the X-ray colours of the sources can be determined with greater accuracy. Hereafter, I will refer to the combined EPIC dataset in these six fields as the *XMM-Newton* Deep One Degree Survey (XMM-DODS).

I have extended the Monte Carlo X-ray colour analysis scheme described in chapter 3 to investigate how well a number of AGN population models reproduce the colour distribution and source counts of the XMM-DODS sample. I restrict the analysis to four of the model N_H distributions introduced in earlier chapters; the $(\log N_H)^5$, $(\log N_H)^8$, *GSH01A*, and *U03* models. The predictions of both the Miyaji et al. (2000). and the Ueda et al. (2003) AGN XLF models are tested against the sample.

In section 5.2 I discuss each of the six XMM-DODS fields, and the method that I have used to define the XMM-DODS sample. I then detail how I have extended the Monte Carlo X-ray colour analysis scheme to the entire XMM-DODS dataset. In section 5.4 I present the results of the X-ray colour comparisons, and in 5.5 I compare the source count distributions in the XMM-DODS sample with the predictions of several XLF models. The amplitude of the source density variations between the XMM-DODS fields are investigated in section 5.5.4. Finally, in section 5.6 I discuss the results and their implications for the AGN population.

5.2 Data reduction method for individual *XMM-Newton* fields

Details of the *XMM-Newton* observations in the six XMM-DODS fields are given in table 5.1. The raw data for the LH, CDFN, LBQS 2212, and 1^H fields have been reduced according to the method detailed in section 2.1. For each observation, images and exposure maps are created for each detector, and for each of the four energy bands (0.2–0.5, 0.5–2, 2–5, and 5–10 keV). The 0.5–2 keV EPIC *pn*+MOS images of these fields are displayed in figure 5.1, illustrating the larger spatial coverage of the LH field.

For consistency, I have carried out the iterative background fitting and source searching process in a uniform way for all six of the XMM-DODS fields. The source

searching was carried out using the same set of EMLDETECT parameters, and multi-band detection likelihood threshold as detailed in sections 4.2.2 and 4.3.1. A summary of the *XMM-Newton* data, and the numbers of sources detected in the six XMM-DODS fields is given in table 5.2. I now discuss each of the fields in more detail.

In the following sections I refer to the total "*pn*-thin filter equivalent" exposure time. This is the summed exposure time from the *pn*+MOS1+MOS2 detectors for a particular observation, after periods of high background have been removed, and after correction for the relative (energy-band dependent) throughput of the particular EPIC detector+filter combinations.

5.2.1 The 1^{H} field

The three *XMM-Newton* observations in the 1^{H} field have been designed to match those in the 13^{H} field, as part of a combined deep field project. The observations are centred at $01^{H}45^{m}00^{s} - 04°35'00''$, in a region of fairly low Galactic column density ($N_H = 2.4 \times 10^{20}$ cm^{-2}). As with the 13^{H} field, the 1^{H} field has been covered by a mosaic of 4×30 ks *Chandra* observations. The *XMM-Newton* observations have been tied to the *Chandra* astrometry, which has in turn been tied to deep optical imaging of the field (Gunn et al., 2006). Periods of high background mean that the useful exposure time in the 1^{H} field is reduced to approximately half of the nominal 200 ks (see tables 5.1 and 5.2). However, despite the lower exposure time, a similar number of sources are detected in the 1^{H} field as are detected in the 13^{H} field. The ongoing spectroscopic identification program in the 1^{H} field has identified one of the X-ray sources to be a Galactic star. I have removed this source from the sample. I also remove four extended sources which are likely to be groups or clusters. After these non-AGN sources are removed, there are 216 detections in the 1^{H} sample. Note that an additional long *XMM-Newton* observation has been made of the 1^{H} field, but that no data have been released for this observation.

Table 5.1: Summary of *XMM-Newton* observations in the 13^H, LH, CDFS, CDFN, LBQS 2212 and 1^H fields, showing the observation ID numbers, dates, nominal-pointings, filters and exposure times for each EPIC instrument. Exposure times are given in kiloseconds for the pn and for each of the two MOS detectors, and indicate the length of good time remaining after removing periods of high background.

Field	Observation	date	RA	Dec	PA	MOS1		MOS2		pn	
					deg	exp (ks)	Filter	exp (ks)	Filter	exp (ks)	Filter
13hr	0109660801	2001-06-12	13:34:37.1	+37:53:02.6	303	43.4	Thin	46.2	Med	35.6	Thin
13hr	0109660901	2001-06-22	13:34:37.2	+37:53:01.9	303	14.2	Med	12.1	Thin	7.4	Thin
13hr	0109661001	2001-06-23	13:34:37.1	+37:53:01.6	303	59.6	Thin	60.6	Thin	53.0	Thin
13hr	TOTAL	-	-	-	-	117.2	-	118.9	-	96.0	-
LH	0123700101	2000-04-27	10:52:41.4	+57:27:06.9	313	32.0	Thin	33.1	Thick	31.9	Thin
LH	0123700201	2000-04-29	10:52:42.9	+57:27:16.7	-48	35.3	Thick	31.5	Thin	51.9	Thin
LH	0123700401	2000-05-02	10:52:42.3	+57:27:04.1	309	10.0	Thin	13.5	Thick	12.4	Thin
LH	0123700901	2000-05-05	10:52:41.5	+57:27:15.4	307	1.5	Thin	2.4	Thick	2.2	Thin
LH	0123701001	2000-05-19	10:52:40.8	+57:27:17.8	-52	23.5	Thin	34.2	Thick	22.8	Thin
LH	0022740201	2001-10-27	10:52:46.2	+57:30:26.9	137	17.0	Med	16.2	Med	16.8	Med
LH	0022740301	2001-11-04	10:52:44.1	+57:30:26.6	128	30.4	Med	30.2	Med	24.6	Med
LH	0147510101	2002-10-15	10:51:05.9	+57:29:34.0	132	64.4	Med	68.2	Med	38.4	Med
LH	0147510801	2002-10-17	10:51:29.9	+57:29:50.0	132	28.1	Med	32.7	Med	15.6	Med
LH	0147510901	2002-10-19	10:52:44.9	+57:30:25.3	132	32.7	Med	35.3	Med	18.6	Med
LH	0147511001	2002-10-21	10:52:10.4	+57:30:13.3	132	64.1	Med	66.9	Med	55.4	Med
LH	0147511101	2002-10-23	10:53:20.2	+57:30:50.8	132	39.8	Med	42.8	Med	21.2	Med
LH	0147511201	2002-10-25	10:54:00.5	+57:31:11.0	132	17.7	Med	19.6	Med	14.9	Med
LH	0147511301	2002-10-27	10:54:31.8	+57:31:30.7	132	17.8	Med	18.5	Med	12.4	Med
LH	0147511601	2002-11-27	10:52:38.5	+57:30:24.6	104	105.2	Med	106.6	Med	93.4	Med
LH	0147511701	2002-12-04	10:52:36.4	+57:30:04.1	104	95.9	Med	96.7	Med	85.8	Med
LH	0147511801	2002-12-06	10:52:40.9	+57:30:44.3	104	81.9	Med	83.1	Med	71.2	Med
LH	TOTAL	-	-	-	-	697.4	-	731.4	-	589.6	-

Table 5.1 continued.

Field	Observation	date	RA	Dec	PA	MOS1		MOS2		pn	
					deg	exp (ks)	Filter	exp (ks)	Filter	exp (ks)	Filter
CDFS	0108060401	2001-07-27	03:32:19.5	-27:47:33.8	59	†20.2	Thin	20.2	Thin	13.4	Thin
CDFS	0108060501	2001-07-28	03:32:22.0	-27:47:53.2	59	†39.6	Thin	40.7	Thin	31.7	Thin
CDFS	0108060601	2002-01-13	03:32:35.0	-27:49:31.0	239	†47.6	Thin	47.6	Thin	42.0	Thin
CDFS	0108060701	2002-01-14	03:32:33.8	-27:49:18.7	239	†73.5	Thin	73.1	Thin	67.9	Thin
CDFS	0108061801	2002-01-16	03:32:34.9	-27:49:10.5	239	†54.8	Thin	54.8	Thin	36.5	Thin
CDFS	0108061901	2002-01-17	03:32:35.0	-27:48:48.9	239	†42.6	Thin	42.6	Thin	38.9	Thin
CDFS	0108062101	2002-01-20	03:32:36.3	-27:49:00.5	239	†43.2	Thin	43.5	Thin	40.6	Thin
CDFS	0108062301	2002-01-23	03:32:35.2	-27:48:53.6	239	†74.3	Thin	72.8	Thin	69.3	Thin
CDFS	TOTAL	-	-	-	-	395.8	-	395.2	-	340.2	-
CDFN	0111550101	2001-05-18	12:36:47.8	+62:11:15.2	312	45.4	Thin	45.4	Thin	42.4	Thin
CDFN	0111550201	2001-05-18	12:36:47.8	+62:11:30.4	312	45.5	Thin	45.5	Thin	41.1	Thin
CDFN	0111550301	2001-05-27	12:36:50.8	+62:11:30.0	312	45.4	Thin	45.5	Thin	37.6	Thin
CDFN	0111550401	2001-06-01	12:36:54.8	+62:11:47.5	312	94.4	Thin	94.4	Thin	91.8	Thin
CDFN	0162160201	2003-11-24	12:36:54.3	+62:14:56.7	134	14.9	Thin	14.9	Thin	11.5	Thin
CDFN	0162160401	2003-12-06	12:36:51.6	+62:14:57.4	123	10.6	Thin	10.6	Thin	8.9	Thin
CDFN	0162160601	2003-12-14	12:36:50.0	+62:14:56.4	116	12.9	Thin	12.9	Thin	11.3	Thin
CDFN	TOTAL	-	-	-	-	269.1	-	269.2	-	244.6	-
LBQS2212	0106660101	2000-11-17	22:15:37.9	-17:44:53.5	250	57.1	Thin	57.1	Thin	54.4	Thin
LBQS2212	0106660201	2000-11-18	22:15:37.9	-17:44:56.4	250	44.1	Thin	43.4	Thin	30.4	Thin
LBQS2212	0106660401	2001-11-17	22:15:37.7	-17:44:57.1	250	22.0	Thin	20.8	Thin	5.5	Thin
LBQS2212	0106660501	2001-11-17	22:15:37.7	-17:44:57.1	250	7.8	Thin	7.9	Thin	5.5	Thin
LBQS2212	0106660601	2001-11-17	22:15:37.2	-17:44:55.3	250	†84.9	Thin	85.8	Thin	60.7	Thin
LBQS2212	TOTAL	-	-	-	-	216.0	-	215.1	-	156.4	-
1hr	0109661101	2002-12-25	01:45:32.6	-04:35:40.5	250	47.7	Thin	47.8	Thin	42.0	Thin
1hr	0109661201	2002-07-16	01:45:21.5	-04:33:40.3	69	47.1	Thin	47.5	Thin	41.8	Thin
1hr	0109661401	2002-07-17	01:45:21.5	-04:33:39.2	69	3.0	Thin	3.3	Thin	2.7	Thin
1hr	TOTAL	-	-	-	-	97.8	-	98.6	-	86.6	-

† 0.2-0.5 keV data excluded from MOS1 CCD chip #5 because of enhanced background

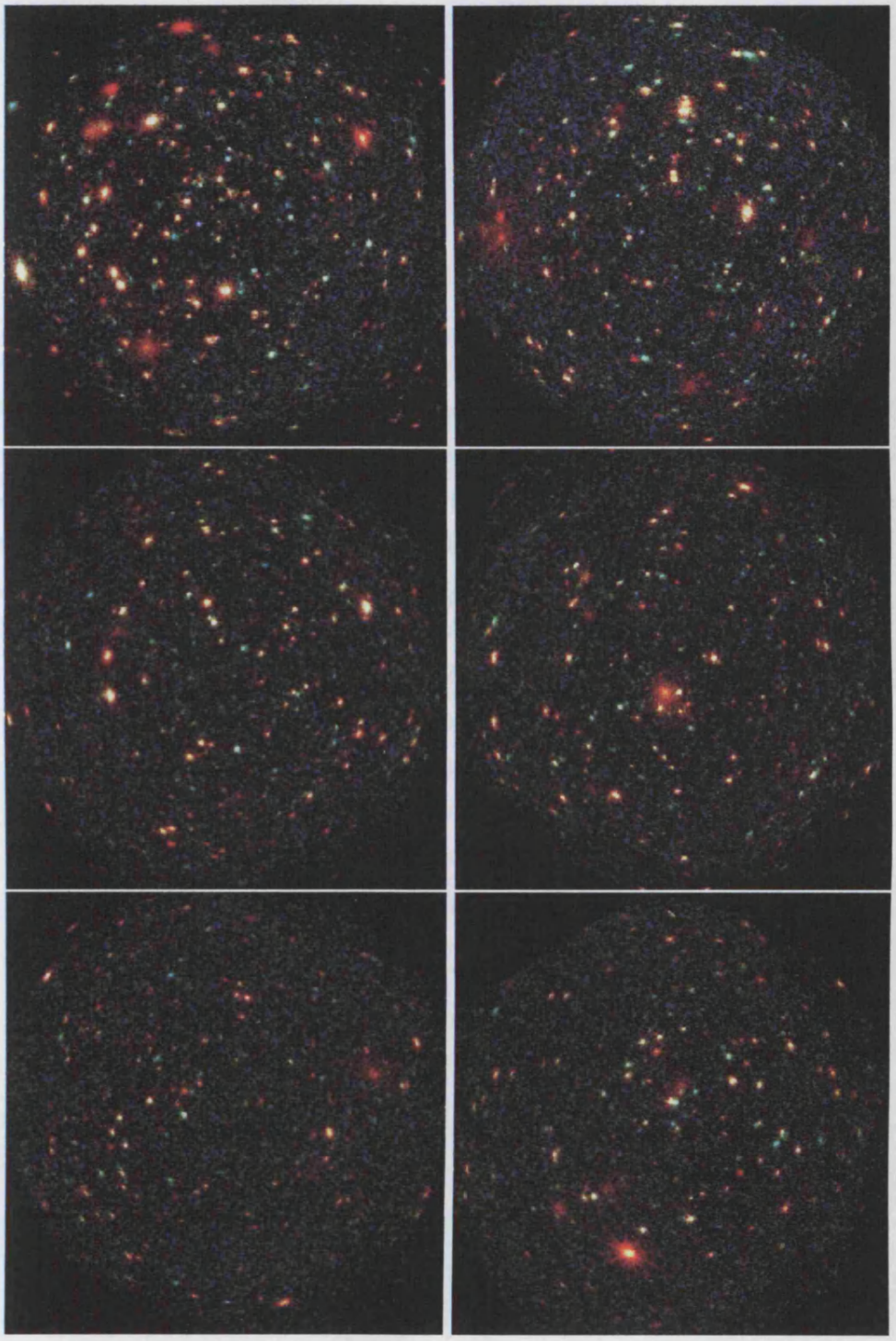

Figure 5.1: Pseudo colour images of the six XMM-DODS fields. The fields are as follows: LH (centre region only, top left), CDFS (top right), CDFN (middle left), LBQS 2212 (middle right), 1^{H} (bottom-left), and 13^{H} (bottom-right). The images show the combined MOS+pn data in the 0.5–2 keV (red), 2–5 keV (green) and 5–10 keV (blue) energy bands, smoothed, background subtracted, and displayed on a square-root scale.

Table 5.2: High level summary of the *XMM-Newton* data in the 13^{H}, LH, CDFS, CDFN, LBQS 2212 and 1^{H} fields. N_H is the Galactic column in the direction of each field. Exposure is the peak thin filter *pn*-equivalent exposure time in the 0.5–2 keV band (see 5.2). The area is the sky area covered with at least 50 ks thin-filter *pn*-equivalent exposure time. Total is the number of sources meeting the detection criteria. The numbers of Galactic stars and groups/clusters identified in each field are also indicated.

Field	Gal. N_H (10^{20} cm^{-2})	exp. time (ks)	Area (deg^2)	Total	Stars	Groups/ Clusters	Other
13^{H}	0.6	157.4	0.145	219	6	3	210
LH	0.6	823.9	0.249	558	4	11	543
CDFS	0.8	537.4	0.182	335	15	15	308
CDFN	1.5	380.1	0.187	286	7	3	276
LBQS 2212	2.4	265.0	0.167	269	6	3	260
1^{H}	2.8	134.9	0.132	221	1	4	216
TOTAL	-	2298.7	1.076	1888	39	39	1810

5.2.2 The 13^{H} field revisited

The four Galactic stars and four extended sources identified in the 13^{H} field (see section 3.3.5) were removed from the sample. After applying the full band detection likelihood threshold that was used in section 4.2.2, and after removing sources which have poorly constrained positions or extension, 210 sources remain.

5.2.3 The CDFN field

The *Chandra* Deep Field-North (CDFN) has been observed with *XMM-Newton* on seven occasions with a total on-axis (*pn*-thin filter equivalent) exposure time (see 5.2) of 380 ks. The *XMM-Newton* observations are centred approximately on $12^{H}36^{m}50^{s} + 62°11'30''$, a part of the sky with fairly low Galactic column density (1.5×10^{20} cm^{-2}). I have tied the *XMM-Newton* astrometric frame to the positions of *Chandra* sources in the 2Ms Alexander et al. (2003) catalogue, which in turn is tied to the optical frame. Using the published follow up observations of the *Chandra* sources (Bauer et al., 2004) I find that there are six stars detected in the *XMM-Newton* observations. I remove these from the sample. I have manually identified three detections which are obviously extended in the *XMM-Newton* images: these are likely to be galaxy groups/clusters and so they are removed from the sample. After removing these non-AGN sources, the final sample of *XMM-Newton* sources in the CDFN contains 276 sources.

5.2.4 The Lockman Hole field

The Lockman Hole (LH) field (RA = $10^{H}52^{m}40^{s}$ Dec=$+57°29'22''$) has been the target for the longest *XMM-Newton* observations to date. The LH is a region of extremely low Galactic column density (6×10^{19} cm^{-2}). There are 17 separate *XMM-Newton* observations in the LH, and most of the exposures have centres within $\sim 2'$ of the nominal pointing. However six of the observations have centres arranged along a line of approximately equal declination, spanning $\sim 30'$ in the RA direction. The

total sky coverage of the LH field is therefore approximately twice that of a single *XMM-Newton* pointing. The LH images and exposure maps have been produced in the same way as described in section 2.1. The reduction of the LH *XMM-Newton* dataset was carried out by Mat Page, and the images have a slightly different pixel scale to the other fields; 4.0″ pixel^{-1} compared to the 4.35″ pixel^{-1} used in the 13^{H}, CDFS, CDFN, LBQS 2212 and 1^{H} fields.

There have been a wealth of multi-wavelength observations in the LH, ranging from radio to the hard X-rays, including a spectroscopic identification program for the X-ray sources. A number of X-ray sources detected in the *ROSAT* and *ASCA* observations of the LH have been identified as stars (Lehmann et al., 2001; Ishisaki et al., 2001). Four of these are detected in the *XMM-Newton* observations, and so I remove them from the sample. I have manually examined the *XMM-Newton* images for regions of extended emission, and remove eleven sources because they are likely to be due to emission from groups/clusters of galaxies. After removing the non-AGN sources, the LH sample contains a total of 543 sources.

Point spread function correction in the LH

The range of pointing centres in the LH raises issues for the source detection process. The EMLDETECT source detection task uses a model of the *XMM-Newton* PSF to detect and measure the properties of candidate sources in the EPIC images. The PSF at any particular position in the final composite image is a summation of the off-axis dependent PSF at that location in each separate observation. The precise shape of the PSF at any location (and energy) is dependent on the contribution to the total exposure time from each of separate observation. The EMLDETECT detect routine uses a PSF model which is dependent on the distance from the nominal centre of the composite image. Of particular importance here, it uses the PSF model to correct the measured count rates (and hence fluxes) for the fraction of the counts from a source that are expected to fall outside the source extraction aperture. The effects of this composite PSF are reproduced in the Monte Carlo

simulation process because each observation is modelled separately, then summed to create a final image. The composite PSF is therefore expected to introduce systematic differences, dependent on both energy and position in the FOV, to the fluxes measured for sources in the LH. In addition, because the vignetting and PSF of EPIC are energy dependent, the composite PSF will affect the measurement of X-ray colours.

In order to quantify and correct for the composite PSF, I have carried out a set of simulations of the LH field. These simulations use a modified form of the Monte Carlo scheme that was described in chapters 2 and 3. Because I wish to investigate only the effects of the composite PSF, I generate simulated EPIC images for an input population of bright, well separated sources, and do not include the effects of photon counting noise. To remove the effects of confusion, the input sources are positioned randomly in the field, but are not permitted to lie within 40″ of each other. As before, a separate simulated image is created for each observation, energy band, and detector (MOS1, MOS2 and *pn*). These images are summed to produce a single composite image for each of the four energy bands. These images are source searched separately in each energy band using EBOXDETECT and EMLDETECT. The output detections in each energy band are matched to input counterparts using a 10″ radius. The ratio of output to input flux (F/S) is calculated for each output detection. The simulation process is repeated for the equivalent of 100 fields, resulting in over 100000 output sources in each of the four energy bands.

I then divide the LH field into a regular grid, each cell of which is 30 pixels ($2' \times 2'$) in size. The median of the F/S ratio of the simulated output sources is measured for each spatial cell and in each energy band. These median values represent the inverse of the count rate correction that should be applied to the real sources in the LH.

In figure 5.2 I show a schematic of the calculated correction factors. For the majority of the cells, the corrections are less than a few percent in any energy band. The PSF differs most from the model used by EMLDETECT at the East and West

"wings" of the field, and it can be seen that this is where the F/S ratio is furthest from unity. This is to be expected; EMLDETECT expects the photons detected in these regions of the composite image to have arrived in the EPIC detectors at high off axis angles, and so models the sources with a broad PSF. However, in actuality, the bulk of the counts in these regions arrived during the short flanking observations, and so were detected closer to the optical axis of the telescope, where the PSF is narrower. In the centre of the LH field, the F/S ratios are less than unity, i.e. the true PSF in these regions is broader than the PSF model of EMLDETECT. This is because some of the source counts in the central regions were detected at high off axis angles during the flanking observations.

I have applied the PSF correction factors to the multi-band countrate, flux and hardness ratio measurements for all the *XMM-Newton* sources in the LH field, according to their position in the FOV. Later, when I carry out multi-band simulations of the fields, I also apply these PSF mismatch corrections to the output simulated sources.

5.2.5 The LBQS 2212 field

The optically bright broad absorption line quasar LBQS 2212–1759 (hereafter LBQS-2212), was observed by *XMM-Newton* on five occasions for a total exposure time of ~220 ks. Somewhat remarkably, despite the NUV/optical brightness of LBQS 2212 ($B \sim 18$), the quasar is not actually detected in the EPIC observations (Clavel, Schartel & Tomas, 2006). The LBQS 2212 field is not strictly speaking, a "blank field" observation, because for example, the target quasar (at $z = 2.217$) could be a "signpost" for a particularly high density part of the universe at this redshift. However, the bias that might arise from using this field is unlikely to be large, especially in a study such as this in which several different fields are analysed. As an added precaution, when analysing the *XMM-Newton* data in this field, I mask out a central $\sim 3'$ region. This is partly to exclude any X-ray sources which may be

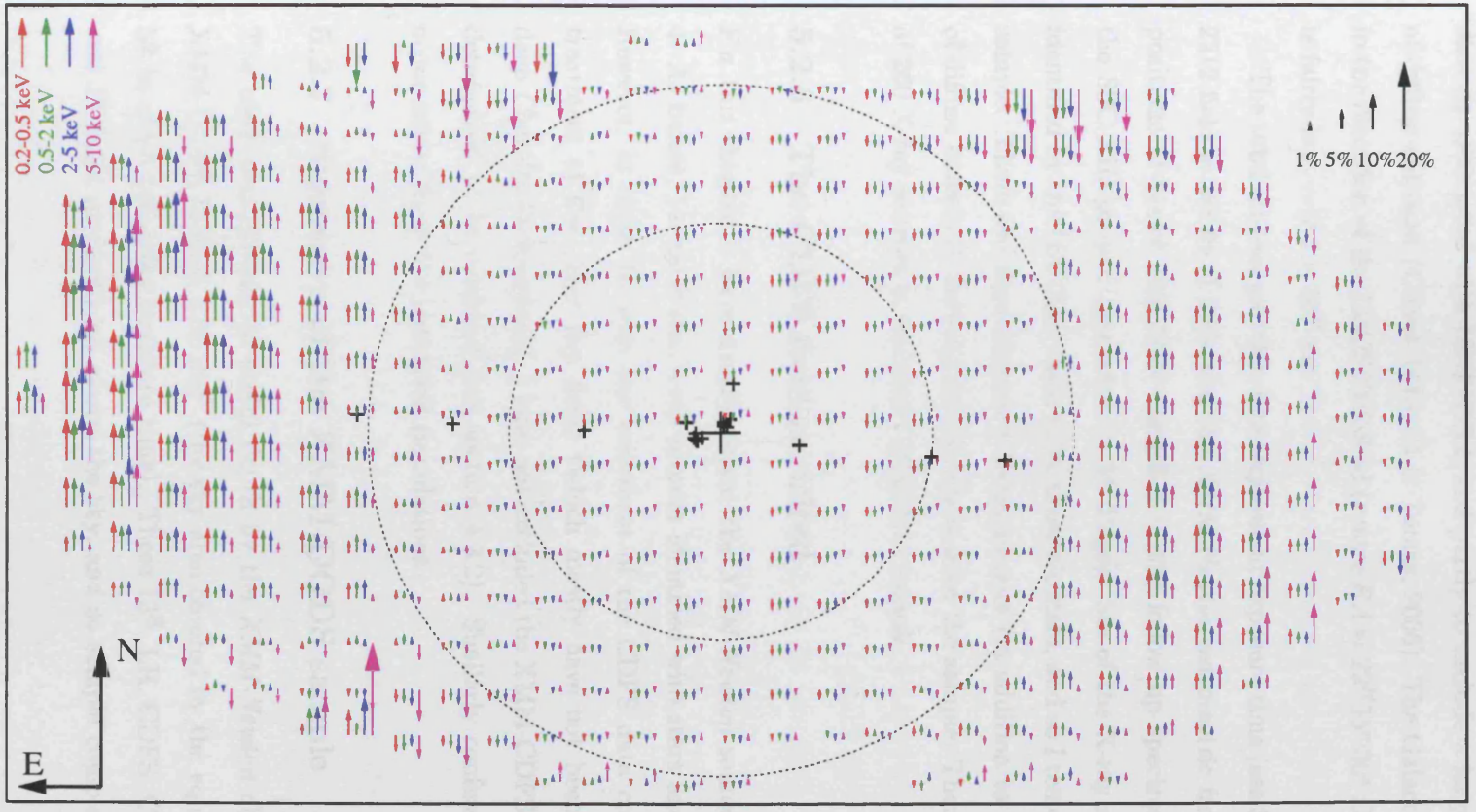

Figure 5.2: PSF flux mismatches in the LH caused by the multiple pointing centres of the LH observations. The arrows show the fractional flux mismatch $((F - S)/S)$ for each $2' \times 2'$ spatial bin, and for each of the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands. Upward pointing arrows show that the output flux (F) is greater than the input flux (S), and vice versa for downward pointing arrows. The large cross shows the exposure-weighted mean optical axis, and the small crosses show the optical axes of the 17 individual observations. The circles are drawn at off axis angles of $9'$ and $15'$ from the mean pointing.

directly associated with LBQS 2212, and partly to exclude a large extended region of diffuse emission (Clavel, Schartel & Tomas, 2006). The Galactic column density in the direction of the LBQS 2212 field (centre $RA = 22^H15^m30^s$ $Dec = -17°44'18''$) is fairly low, $\sim 2.4 \times 10^{20}$ cm^{-2}.

The total on-axis *pn*+thin filter equivalent exposure time (see 5.2) in the LBQS-2212 field is 258 ks. I have tied the *XMM-Newton* astrometric frame to the optical positions of broad line AGN identified in the follow up spectroscopy program of the SSC-XID project (provided by M. Page). Six of the X-ray sources have been identified by the SSC-XID team to be Galactic stars, and so I remove these from the sample. There are three extended X-ray sources (in addition to the central region of diffuse emission), and so I remove these from the sample. There are then a total of 260 X-ray sources in the final LBQS 2212 sample.

5.2.6 The CDFS field revisited

For this chapter, I have simply reused the *XMM-Newton* sourcelist from chapter 4. As before, I remove the X-ray sources identified with stars and clusters/groups. However, in order to keep the treatment of the CDFS data consistent with the treatment of the other five fields (which mostly have not been covered by very deep *Chandra* observations), I have not excluded the XMM-CDFS sources that were determined to be confused (see section 4.3.2). Similarly confused detections will occur naturally in the simulated populations.

5.2.7 Summary of the XMM-DODS sample

The total area covered to useful depth by the *XMM-Newton* observations in the XMM-DODS fields is 1.08 deg^2 (the sky area covered by the equivalent of at least 50 ks of *pn* thin-filter exposure time). These 13^H, LH, CDFS, CDFN, LBQS 2212 and 1^H fields are scattered about the sky, and so sample completely unconnected

parts of the Universe. The total number of XMM-DODS sources, after the multi-band detection threshold criteria have been applied, and after non-AGN objects have been removed, is 1810. This is more than eight times the sample size used in chapter 3. The optically identified non-AGN sources have been removed from the XMM-DODS sample as well as clearly extended X-ray sources. The total number of sources removed is 78 (see table 5.2). However, because the optical identification programs for these fields are incomplete, some small fraction of non-AGN sources are likely to remain. The fraction of non-AGN sources found in the XMM-CDFS sample can be used to estimate the number of non-AGN sources remaining in the XMM-DODS sample. Of the 335 X-ray detections in the XMM-CDFS sample, there are 15 sources identified as stars, and 15 sources which are likely to be galaxy groups or clusters. Taking this fraction ($\sim$9%) to be representative, then one might expect that $\sim$170 out of the 1888 X-ray detections in the XMM-DODS sample to be non-AGN sources. Therefore, taking into account the number of non-AGN sources already removed from the XMM-DODS sample, I estimate the remaining fraction of non-AGN sources to be $\leq$ 5%. Note that this is an upper limit because many of the stars and groups/clusters in the XMM-CDFS sample were detected at very faint fluxes (a few times 10^{-16} erg s^{-1} cm^{-2}). Similarly faint objects would not be detected in the somewhat shallower *XMM-Newton* observations of the CDFN, LBQS 2212, $13^{\rm H}$ and $1^{\rm H}$ fields.

5.3 Modelling the XMM-DODS sample

I have adapted the Monte Carlo simulation process (used in chapters 3 and 4) in order to compare the observed source count distributions and X-ray colours in the XMM-DODS sample, to the predictions of a number of AGN population models.

Building upon the results of the two previous chapters, and in order to reduce the processing time requirement, I have limited the scope of this chapter to testing the predictions of the $(\log N_H)^5$, $(\log N_H)^8$ and *GSH01A*, and *U03* N_H models. I also

wish to investigate the effect of different model XLFs on the predicted distribution of X-ray colours and the predicted source count distribution. Therefore, the AGN populations are modelled using both the LDDE1 XLF model of Miyaji et al. (2000) as well as the XLF model of Ueda et al. (2003), which predict markedly different distributions of AGN in redshift/luminosity space (see section 4.6.3).

I have generated simulated populations following the method of section 4.5.7. One hundred fields of simulated images are generated, background fitted, and source searched for each of the six XMM-DODS fields, for each of the four N_H models, and for both of the XLF models; a total of 4800 repetitions of my simulation process. The workload was distributed between more than a dozen workstations in order to reduce the absolute length of time taken to carry out the processing. In a similar way to earlier chapters, the absolute normalisations of the model AGN populations have been iteratively adjusted to match the observed integral source counts. However, for the purposes of this chapter, the normalisation for each combination of N_H model + XLF model is iteratively adjusted so that at a 0.5–2 keV flux of 2×10^{-15} erg s^{-1} cm^{-2}, the simulated output source counts, summed over all six XMM-DODS fields, are matched to within 10% of the number of real source counts summed over all six XMM-DODS fields. The small residual fractional differences between the observed and simulated output source counts are noted, and are taken into account later.

As discussed in earlier chapters, because of the rescaling of the XLF normalisation some caution must be taken when interpreting the different predictions made by the N_H models. For example, when comparing the model source count distributions with those measured in the XMM-DODS sample, the absolute normalisation can obviously not be considered a free variable.

5.4 X-ray colour distribution in the XMM-DODS sample

In this chapter I adopt "flux ratios" as a measure of X-ray colour rather than the hardness ratios (based on count rates) used earlier. This different approach is necessary because of the range of Galactic column densities ($0.6 - 2.8 \times 10^{20}$ cm^{-2}) in the directions of the six XMM-DODS fields. The flux ratios, FR, are defined in essentially the same way as the hardness ratios used before,

$$FR1 = \frac{S_{0.5-2} - S_{0.2-0.5}}{S_{0.5-2} + S_{0.2-0.5}} \tag{5.1}$$

$$FR2 = \frac{S_{2-5} - S_{0.5-2}}{S_{2-5} + S_{0.5-2}} \tag{5.2}$$

$$FR3 = \frac{S_{5-10} - S_{2-5}}{S_{5-10} + S_{2-5}} \tag{5.3}$$

where $S_{0.2-0.5}$, $S_{0.5-2}$, S_{2-5}, and S_{5-10} are the fluxes measured in the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands respectively. The corresponding measurement errors are denoted by σ_{FR1}, σ_{FR2}, and σ_{FR3} and are calculated from the errors on the fluxes thus,

$$\sigma_{FR1} = 2\frac{\sqrt{(\sigma_{S0.5-2}\, S_{0.2-0.5})^2 + (\sigma_{S0.2-0.5}\, S_{0.5-2})^2}}{(S_{0.5-2} + S_{0.2-0.5})^2} \tag{5.4}$$

$$\sigma_{FR2} = 2\frac{\sqrt{(\sigma_{S2-5}\, S_{0.5-2})^2 + (\sigma_{S0.5-2}\, S_{2-5})^2}}{(S_{2-5} + S_{0.5-2})^2} \tag{5.5}$$

$$\sigma_{FR3} = 2\frac{\sqrt{(\sigma_{S5-10}\, S_{2-5})^2 + (\sigma_{S2-5}\, S_{5-10})^2}}{(S_{5-10} + S_{2-5})^2}. \tag{5.6}$$

For the cases where a source has zero flux in two consecutive energy bands, the corresponding FR measurement is set to be 0.0 ± 1.0.

It should be noted that (as throughout this thesis) the countrate-to-flux conversions for the XMM-DODS sources are calculated assuming a power law spectral model, with a slope $\Gamma = 1.7$. For sources having spectral shapes differing from

this assumed spectral form, the countrate-to-flux conversion will in general over- or underestimate the "true" flux and "true" flux ratios. The absolute values of the flux ratios defined above are thus also dependent on the choice of spectral model. However, for the simulated source populations examined later on I use the "output" source fluxes that are determined from countrates measured in simulated images, in exactly the same way as for the XMM-DODS sources. Therefore when I make comparison between the fluxes and flux ratios distributions of the XMM-DODS sources with those predicted by simulated populations, I do not expect the assumed spectral model to affect my results greatly.

I present the distribution of the XMM-DODS sample in flux ratio space in figure 5.3. The bulk of the XMM-DODS sources have flux ratios consistent with an "absorbed powerlaw plus reflection" spectral model (as introduced in earlier chapters), illustrated by the track shown in figure 5.3. The contours show that the peak of the $FR1/FR2$ distribution is broadly in agreement with an AGN population with range of absorptions and redshifts, with a mean intrinsic spectral slope of $\bar{\Gamma} \sim 1.9$. However in $FR2/FR3$ space the pattern is somewhat different. The peak of the measured distribution is offset to higher values of $FR3$ than would be expected for the canonical AGN spectral slope of $\Gamma \sim 1.9$. This offset could be due to there being large numbers of sources with hard spectral slopes; the track for $\Gamma = 1.4$ is a much closer match to the peak of the distribution in $FR2, FR3$. Alternatively there may be a lack of softer spectrum sources, skewing the distribution somewhat. Also it should be cautioned that because the EMLDETECT detection algorithm does not fit negative fluxes to sources, so one should not necessarily expect the FR distributions to be perfectly symmetric about the "mean track". Therefore, I have carried out a closer examination of the X-ray colour distribution of the XMM-DODS sample in comparison to the predictions of the AGN population models.

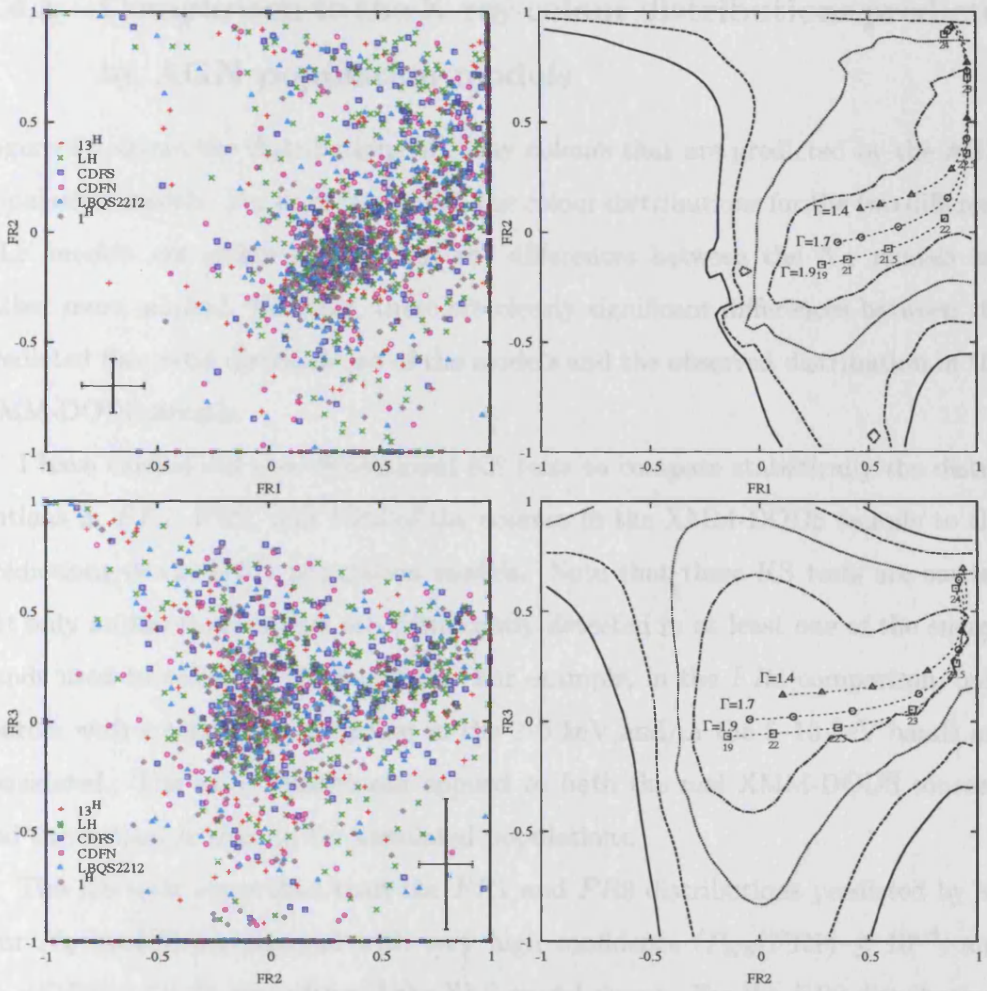

Figure 5.3: Distribution of flux ratios in the XMM-DODS sample, shown as a scatter plot (left panels), and as a smoothed distribution (right panels). Contours are calculated from a smoothed distribution, generated by convolving each of the source positions with a 2-D Gaussian having widths equal to the flux ratio measurement errors. The Gaussian for each source is normalised such that it has a total contribution of unity in the 2-D interval [-1:1][-1:1]. The contours contain 50% (dot-dashed), 75% (dotted), 90% (dashed), and 95% (solid), of the smoothed distribution. The cross in each plot shows the median error on the flux ratios. Also shown are the tracks in colour-colour space for an AGN at $z = 1$ with an absorbed power law plus reflection spectral model, and for three different *intrinsic* spectral slopes.

5.4.1 Comparison to the X-ray colour distributions predicted by AGN population models

Figure 5.4 shows the distributions of X-ray colours that are predicted by the AGN population models. For each N_H model, the colour distributions for the two different XLF models are rather similar, but the differences between the N_H models are rather more marked. However, there are clearly significant differences between the predicted flux ratio distributions of the models and the observed distribution in the XMM-DODS sample.

I have carried out one dimensional KS tests to compare statistically the distributions in $FR1$, $FR2$, and $FR3$ of the sources in the XMM-DODS sample to the predictions of the AGN population models. Note that these KS tests are carried out only on the sources that are significantly detected in at least one of the energy bands used to calculate the flux ratio. For example, in the $FR3$ comparison, only sources with a significant detection in the 2–5 keV and/or the 5–10 keV bands are considered. The same criteria are applied to both the real XMM-DODS sources, and the output sources in the simulated populations.

The KS tests show that both the $FR1$ and $FR3$ distributions predicted by all four N_H models are rejected with very high confidence ($P_{KS}(FR1) \leq 10^{-8}$, and $P_{KS}(FR3) \leq 10^{-9}$), regardless of the XLF model chosen. For the $FR2$ distribution, the $(\log N_H)^5$ and *U03* models are strongly rejected ($P_{KS}(FR2) \leq 6 \times 10^{-7}$, and $P_{KS}(FR2) \leq 10^{-13}$ respectively), when combined with either XLF model. In contrast, and when combined with the XLF of Ueda et al. (2003), the distributions of $FR2$ predicted by the $(\log N_H)^8$ and *GSH01A* N_H models are acceptable matches to the $FR2$ distribution in the XMM-DODS sample at the 1% and 0.1% levels respectively. When these two N_H models are combined with the LDDE1 XLF model of Miyaji et al. (2000), the $P_{KS}(FR2)$ is reduced significantly. Figure 5.5 shows the 1-dimensional cumulative distributions of $FR1$, $FR2$ and $FR3$ in the XMM-DODS sample, together with the predictions of the AGN population models. This plot is

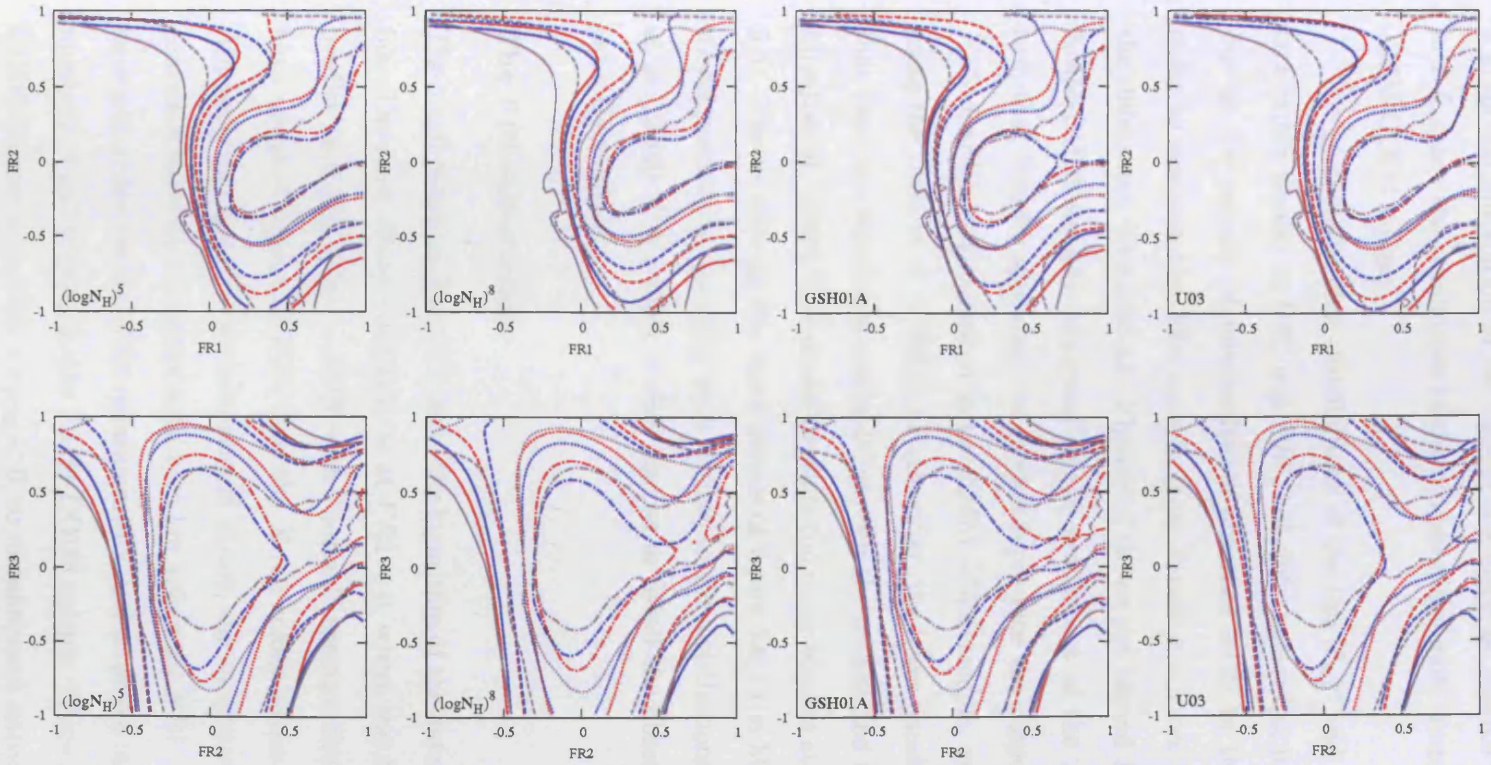

Figure 5.4: Flux-ratio distributions predicted by the four N_H models, when combined with the Ueda et al. (2003) XLF model (blue), and the Miyaji et al. (2000) LDDE1 XLF model (red). The observed distribution in the XMM-DODS sample is shown for comparison (grey). The top row shows $FR1$ *vs* $FR2$, and the bottom row $FR2$ *vs* $FR3$. Contours are generated from a smoothed distribution in the same way as for figure 5.3, and contain 50% (dot-dashed), 75% (dotted), 90% (dashed), and 95% (solid), of the smoothed distribution.

a graphical illustration of the comparison made by the KS tests, and shows where in FR space the predictions of the models differ most from the distribution in the XMM-DODS sample.

The predicted colour distributions of the $(\log N_H)^8$ and *GSH01A* models are in fact rather similar in $FR2$ and $FR3$, but differ somewhat in $FR1$. This is possibly due to the paucity of intermediately absorbed AGN in the *GSH01A* model. In order to examine the differences in more detail, it is more informative to examine the differential distributions. Figure 5.6 shows the binned FR distributions in the XMM-DODS sample compared to the predictions of the N_H models. It can be seen that broadly speaking, the models reproduce the shape of the $FR1$, $FR2$ and $FR3$ distributions measured in the XMM-DODS sample, especially for the models using the Ueda et al. (2003) XLF. However, the lower panels of figure 5.6 illustrate that there are significant residuals between the models and the data. Note that the Miyaji et al. (2000) XLF model predicts too many sources at faint fluxes (see section 5.5). This is why (in the lower panels of figure 5.6) the Miyaji et al. (2000) XLF model predictions are offset with respect to the distributions predicted by the Ueda et al. (2003) XLF model. I now discuss in detail the observed and predicted flux ratio distributions.

The $FR1$ distribution

The small panels of figure 5.6 show the logarithm of the data to model ratios in each bin. The most obvious disparity is at $FR1 < 0$, where the AGN population models predict very few objects. AGN with "normal" spectral slopes are not expected to have negative values of $FR1$ (see the X-ray colour-colour diagnostic diagram in Appendix A). Because of measurement errors, the simulated model populations *do* contain a few output sources with very low values for $FR1$. However, most of these have poorly determined $FR1$ measurements, and these are insufficient to explain the numbers of soft objects in the XMM-DODS sample. Many of the very soft XMM-DODS sources have $FR1 + \sigma_{FR1} < 0$, so measurement uncertainties cannot explain

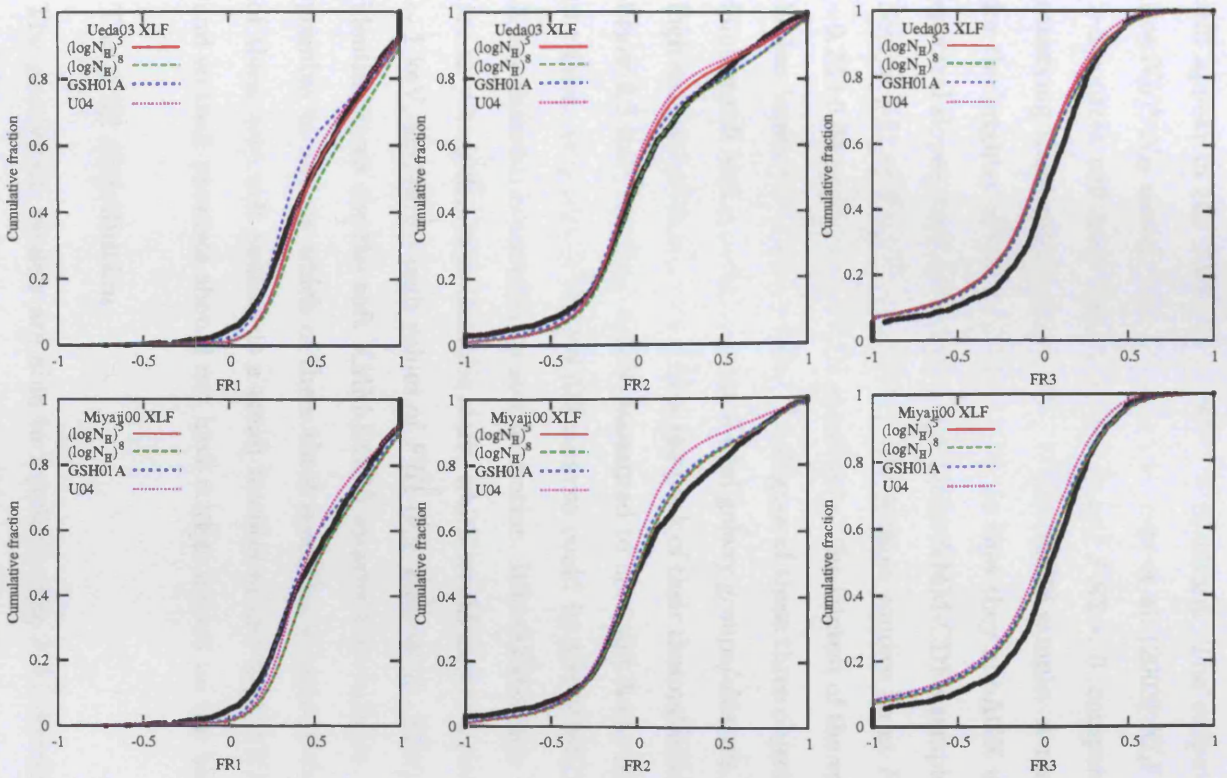

Figure 5.5: The cumulative flux ratio distributions in the XMM-DODS sample (thick lines) compared to the predictions of the AGN population models. The panels show the cumulative distributions for each of the flux ratios $FR1$, $FR2$ and $FR3$. For comparison, I show the predictions of the $(\log N_H)^5$, $(\log N_H)^8$, *GSH01A* and *U03* N_H models for populations generated according to the XLF model of Ueda et al. 2003 (upper row), and the LDDE1 XLF model of Miyaji et al. 2000 (lower row).

their soft colours. Note that the upper left quadrant of the $FR1$ *vs* $FR2$ plot is actually occupied by absorbed sources which have effectively zero flux below 2 keV, and so have essentially undefined measurements of $FR1$. Genuinely soft sources will appear in the $FR1 < 0$, $FR2 < 0$ quadrant. The output simulations for the $(\log N_H)^8$ N_H model, with the XLF of Ueda et al. (2003) XLF model, predict that 2–3 sources will have $FR1 + \sigma_{FR1} < 0$ and $FR2 < 0$, compared to the 25 sources satisfying these criteria in the real XMM-DODS sample. A reasonable explanation for the brighter of the very soft sources is that they are AGN with intrinsically soft spectral slopes, such as the two AGN in the XMM-CDFS sample (section 4.7.3), and Source #52 of Page et al. (2006). These three sources lie at $FR1 = -0.15 \pm 0.01$, -0.23 ± 0.03 and -0.37 ± 0.03 respectively. The faintest of the very soft sources could be less luminous or more distant analogues of these three objects. Alternatively, the faintest soft sources could be unresolved galaxy groups/clusters. If such objects lie at high enough redshift ($z > 1$), then the peak of their thermal emission could be shifted below 0.5 keV, and they will be measured to have soft X-ray colours. Alternatively, the faint, very soft XMM-DODS sources could be low redshift galaxies where the X-ray emission is powered by star formation. It is unlikely that the faint soft objects are as-yet unidentified Galactic stars, which would have the peak of their spectra at ~ 1 keV, and hence high values of $FR1$ (e.g. Page et al., 2006). Only a few optical identifications for the soft XMM-DODS sources are available, and so it is not yet possible to deduce which of these three scenarios is more likely. The total number of these very soft sources is a small fraction of the total XMM-DODS (25/1810), and so their presence should not have a large impact on the findings of this survey.

The $FR2$ distribution

The population models are able to reproduce the $FR2$ distribution of the XMM-DODS sample much better than the $FR1$ distribution, especially the $(\log N_H)^8$ and *GSH01A* models (note the different y-scale to the $FR1$ residual plots in figure 5.6). However there are evidently systematic differences, as demonstrated by the shape of

the residual curves. In particular, the XMM-DODS sample has more sources with $FR2 \sim 0.5$ than predicted by any of the N_H models with the Ueda et al. (2003) XLF. What AGN spectral parameters can produce these AGN with $FR2 \sim 0.5$? Referring to the X-ray colour-colour diagnostic diagram in Appendix A, one can see that to produce more sources with $FR2 \sim 0.5$, the N_H models should contain more AGN with N_H in the range $10^{22.5} - 10^{23.5}$ cm^{-2} (depending on their redshifts). Of the four N_H models, the $(\log N_H)^8$ model contains the largest numbers of input sources in this N_H range, and indeed this model predicts the $FR2$ distribution that is closest to the $FR2$ distribution in the XMM-DODS sample.

The model populations each under-predict, by factors of between 1.5 and 3, the number of XMM-DODS sources in the $FR2 < -0.9$ bin. The $FR2 < -0.9$ sources in the XMM-DODS sample are mostly X-ray faint (median $S_{0.5-2} = 8 \times 10^{-16}$ erg s^{-1} cm^{-2}). It is possible that they are measured to have such low values of $FR2$ because they are detectable in the 0.5–2 keV band, but because of the poorer sensitivity at harder energies, they too faint to be detected above 2 keV. Indeed, the output simulated sources that are detected with $FR2 < -0.9$ in the simulated populations are nearly all faint and unabsorbed. The XLF model of Miyaji et al. (2000) predicts a larger number of faint AGN than the XLF model of Ueda et al. (2003), and so produces a closer match to the number of $FR2 < -0.9$ sources in the XMM-DODS sample. It is also possible that sources detected in this part of flux ratio space are not AGN and have intrinsically soft spectra, e.g. diffuse emission from high-z galaxy groups/clusters, or powerful star forming galaxies.

The $FR3$ distribution

It can be seen in the $FR3$ residuals plots (shown in figure 5.6), that when the N_H models are combined with the Ueda et al. (2003) XLF model, they systematically contain too many sources at $FR3 < 0$ and too few sources at $FR3 > 0$. For model populations generated with the XLF model of Miyaji et al. (2000), the residual curves have a very similar shape, but are offset with respect to the predictions of

the Ueda et al. (2003) XLF model. This is because the Miyaji et al. (2000) XLF model overpredicts the *total* number of sources, so it does quite well at predicting the numbers of $FR3 > 0$ sources, and worse for $FR3 < 0$. The biggest difference is for $FR3 < 0$, where the XMM-DODS sample contains far fewer sources than predicted by any of the population models. This effect can also be seen to a certain extent in the 2-D distribution of figure 5.4. This could indicate that at high energies, the mean spectral slopes of AGN are harder than the mean slope at lower energies. The spectral model I have adopted does include a small component of reflected emission, which has the effect of hardening the spectral slope at restframe energies of $E > 5$ keV. By increasing the contribution of this reflected component one can expect to harden the $FR3$ distribution. However, it is unphysical to make the reflected component much larger than the primary power law component. The observed $FR3$ distribution could therefore be an indication of intrinsic hardening of the primary power law at high energies. This hardening was seen for two of the higher signal-to-noise AGN in the 13^{H} field (Page et al., 2006), which were well fitted with a double powerlaw spectral law model. A definitive answer will require redshifts and a more detailed X-ray spectral fitting of the sources in the XMM-DODS sample.

5.5 Source count distributions in the XMM-DODS sample

Because of the large size, and multi-field nature, an examination of the multi-band source counts in the XMM-DODS sample is expected to be productive, especially to test AGN XLF models. The relative source counts in different X-ray bands can be used as a probe of the absorption distribution. I remind the reader that in chapter 4, although the LDDE1 XLF model of Miyaji et al. (2000) reproduced poorly the source counts in the XMM-CDFS sample, the sample size was insufficient

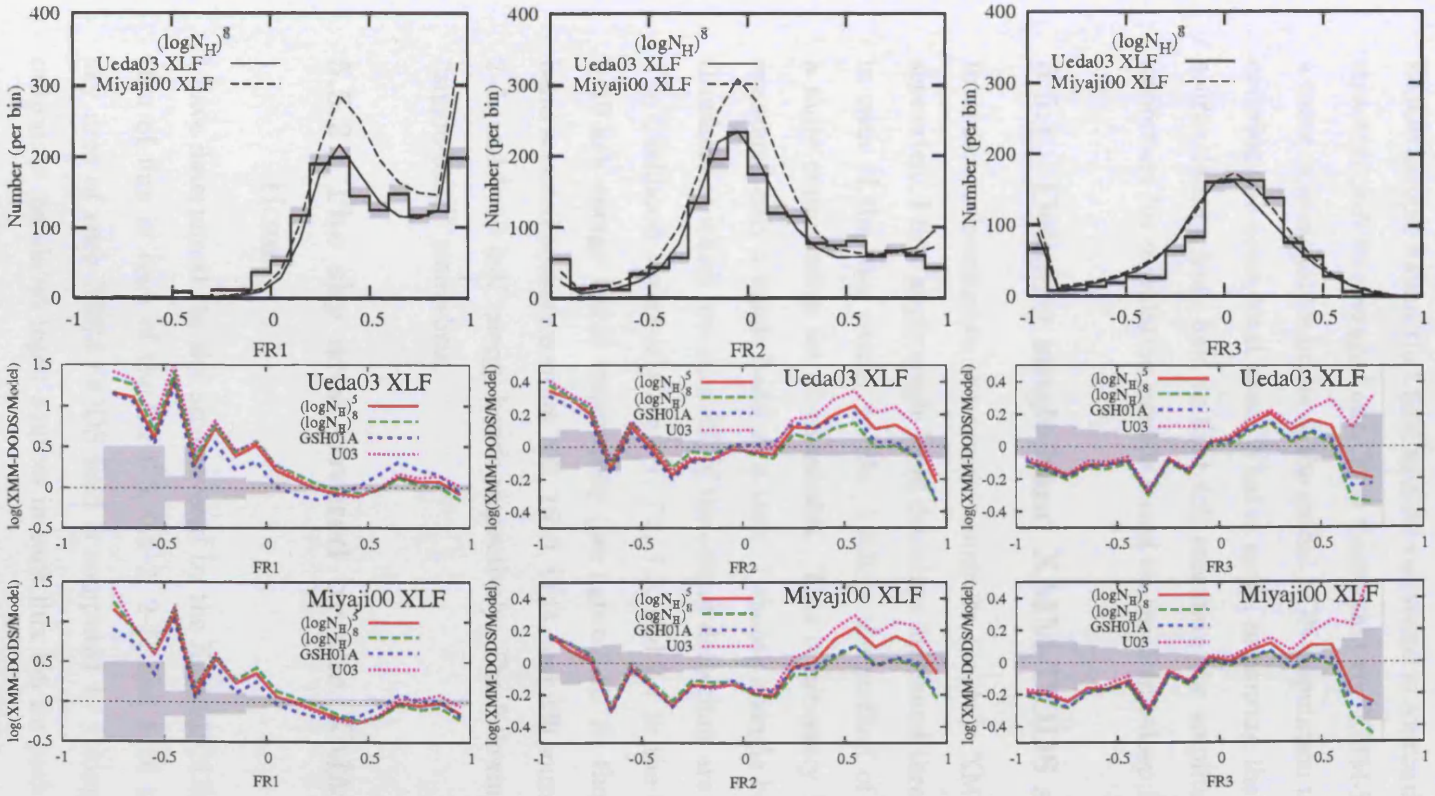

Figure 5.6: The differential flux ratio distributions in the XMM-DODSsample, and the predictions of the AGN population models. The top panels show the distribution observed in the XMM-DODS, for each of $FR1$, $FR2$ and $FR3$, in comparison to the predictions of the $(\log N_H)^8$ N_H model. The grey shaded areas indicate the $(\sqrt{N})$ uncertainty in each bin. The two lower rows of panels show the log of the ratio between the numbers of XMM-DODS objects and the numbers predicted by the model AGN populations. Results are shown for the four N_H models, and the two XLF models. For clarity, the uncertainty on the log of the ratio (shaded area) is displayed at log(XMM-DODS/Model) = 0. Note the different vertical scale for the $FR1$ panels.

to statistically reject the model with certainty. The multi-field nature of the XMM-DODS sample means that field-to-field variations in source density due to large-scale structure will be averaged out. This means that the XMM-DODS sample should be a more representative probe of the global AGN population than a contiguous survey covering the same total area. What is more, measuring the size of the variations in source density from field to field will constrain the amplitude of this effect; this is important for population models based on small-but-deep fields.

5.5.1 Defining single-band XMM-DODS samples

In order to investigate the source counts in the four XMM-DODS energy bands separately, I first apply single band detection likelihood threshold cuts to the sample in each of the four energy bands. I follow the method of section 3.3, but choose a more conservative set of thresholds. This is necessary because the sources are examined only a single band at a time. I choose a single band detection likelihood threshold at which less than 3% of the output detections are expected to be spurious. The likelihood thresholds are 7.1, 7.2, 7.2, and 9.7 in the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands respectively (see figure 3.4). At these thresholds, the single band selected subsets contain 702, 1510, 1144, and 230 sources in the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands respectively. The following sections refer to these "single-band" sourcelists.

5.5.2 The sky area covered by the XMM-DODS observations

I have determined the sky area covered by the XMM-DODS observations as a function of flux in each of the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands. The sky area of each XMM-DODS field is computed by taking the ratio of simulated output to simulated input sources in each flux bin for each energy band, and then multiplying this ratio by the geometric area of the EPIC coverage. For the sky area

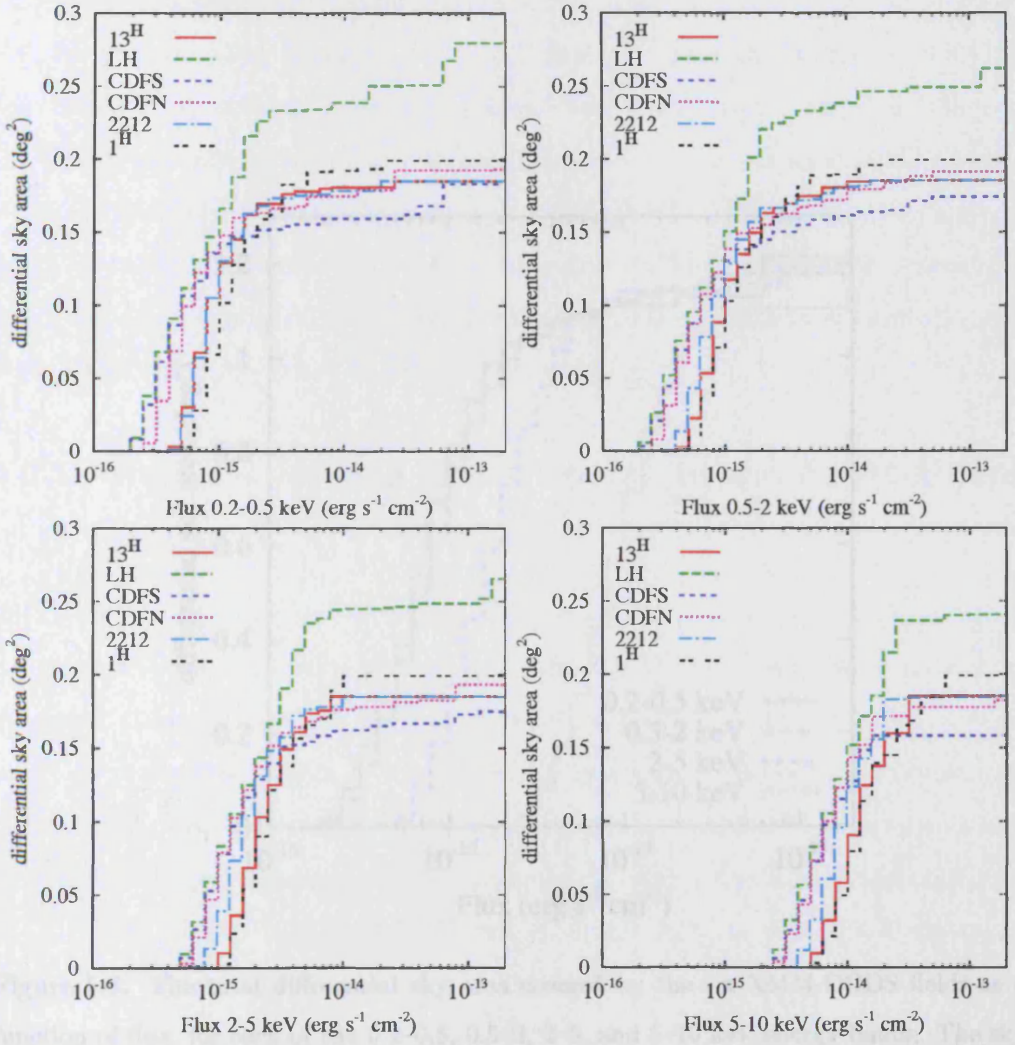

Figure 5.7: The differential sky area covered by the six XMM-DODS fields as a function of flux, for each of the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands. The sky area is calculated from the relative numbers of simulated input and output sources in a number of flux bins, together with the geometric area covered by the EPIC observations in each field. Note that the *XMM-Newton* observations in the Lockman Hole cover a wider area than the other fields, but are deep only in the central $\sim 30'$ region.

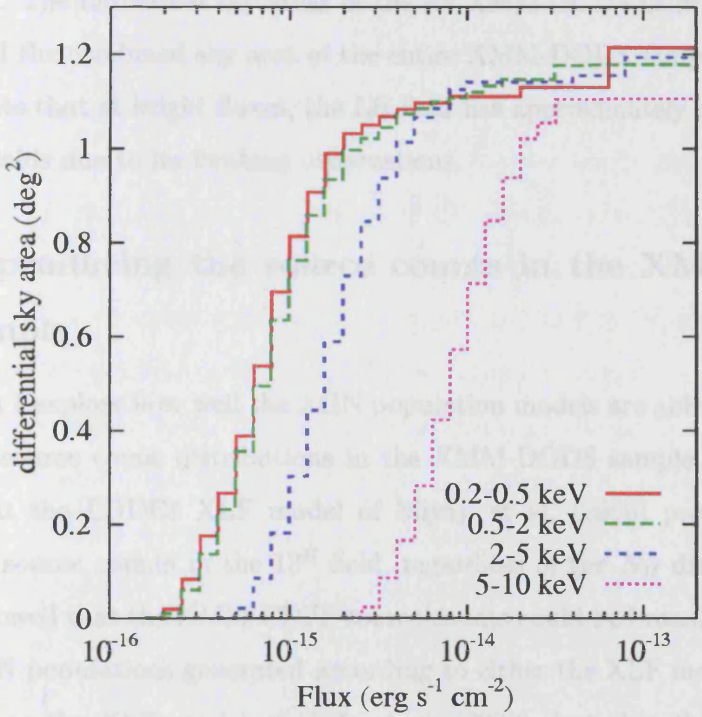

Figure 5.8: The total differential sky area covered by the six XMM-DODS fields as a function of flux, for each of the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands. The sky area is calculated from the relative numbers of simulated input and output sources in a number of flux bins, together with the geometric area covered by the EPIC observations.

calculations, the output simulated sourcelists have been curtailed at the same single band detection likelihood thresholds as the real XMM-DODS sample. In order to maximise the accuracy of this calculation I have used quite narrow bins in flux (0.1 dex width), so in order to have sufficient numbers of simulated sources in each bin, the sky area is calculated using simulated sources from all eight of the AGN population models. The individual sky areas of the six XMM-DODS fields are shown in figure 5.7, and the combined sky area of the entire XMM-DODS sample is shown in figure 5.8. Note that at bright fluxes, the LH field has approximately twice the area of the other fields due to its flanking observations.

5.5.3 Reproducing the source counts in the XMM-DODS sample

In this section I explore how well the AGN population models are able to reproduce the observed source count distributions in the XMM-DODS sample. I showed in chapter 3 that the LDDE1 XLF model of Miyaji et al. (2000) provided a poor match to the source counts in the 13^{H} field, regardless of the N_H distribution. In chapter 4 I showed that the XMM-CDFS source counts could be broadly reproduced by model AGN populations generated according to either the XLF model of Miyaji et al. (2000), or the XLF model of Ueda et al. (2003); but that the Ueda et al. (2003) XLF better reproduced the redshift/luminosity distribution in the XMM-CDFS sample. However, in the latter field, the source counts are dominated by large scale structures, causing "spikes" at $z < 1$ in the redshift distribution. A large sample of AGN is required to differentiate between these two rather different XLF models. In figures 5.9 and 5.10 I present the differential source counts of the XMM-DODS sample in four energy bands, and compare them to the predictions of the XLF models of Ueda et al. (2003), and Miyaji et al. (2000). To test the goodness of the matches, I have carried out KS tests on the source count distributions in each energy band. The results are presented in table 5.3, and confirm the large differences seen

in earlier chapters in the ability of the two XLF models to reproduce the observed source counts.

It is clear from the statistical tests that in the 0.2–0.5, 0.5–2, and 2–5 keV bands, that a model AGN population distributed in redshift/luminosity space according to the LDDE1 XLF model of Miyaji et al. (2000) provides a poor match to the source counts in the XMM-DODS sample, and can be rejected with high certainty. The Miyaji et al. (2000) XLF is rejected for each of the four N_H distribution models examined here. Broadly speaking, the Miyaji et al. (2000) XLF predicts a source population containing too many faint objects and too few bright objects; as noted earlier in sections 3.6.2, and 4.6.2. In the 5–10 keV band, the LDDE1 model does provide a statistically adequate match to the observed source counts, for all of the N_H models. The low numbers of sources detected above 5 keV makes this a relatively weak constraint. More importantly, the 5–10 keV source counts do not reach to the faint fluxes where in the softer bands, the predictions of the LDDE1 model over-predict the numbers of observed sources. So in conclusion, the LDDE1 XLF model of Miyaji et al. (2000) is evidently a poor description of the real distribution of luminosities and redshifts in the AGN population below 5 keV. In particular, it should be highlighted that the LDDE1 model does not reproduce the source count distribution in the energy band in which it is defined (0.5–2 keV).

In contrast, the Ueda et al. (2003) XLF model predicts the observed XMM-DODS source counts rather well. When combined with the $(\log N_H)^5$, $(\log N_H)^8$ and *GSH01A* N_H models, this XLF model is able to reproduce well the source counts in the 0.5–2, 2–5, and 5–10 keV bands over nearly the entire flux range sampled by the XMM-DODS sources (see figure 5.9). The *U03* N_H model provides a statistically poorer description (rejected with $> 99\%$ confidence) of the source counts in the 0.5–2 and 2–5 keV bands. There is suggestion that the models overpredicts the numbers of very bright ($S > 10^{-14}$ erg s^{-1} cm^{-2}) sources in the 2–5 and 5–10 keV bands. However, for this flux regime, rather few sources are expected in the one square degree covered by XMM-DODS. An analysis of a much wider area *XMM-Newton*

Table 5.3: Statistical comparison of the observed and model source count distributions in the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands. $P_{KS}(S)$ is the KS test probability in the given band. The comparisons are made for the real and simulated sources which satisfy the detection likelihood criteria for the band in question.

N_H model	XLF model	$P_{KS}(S_{0.2-0.5})$	$P_{KS}(S_{0.5-2})$	$P_{KS}(S_{2-5})$	$P_{KS}(S_{5-10})$
$(\log N_H)^5$	Ueda 2003	0.05	0.30	0.06	0.92
$(\log N_H)^8$	Ueda 2003	0.02	0.65	0.14	0.84
GSH01A	Ueda 2003	0.48	0.25	0.09	0.85
U03	Ueda 2003	0.26	0.004	0.006	0.77
$(\log N_H)^5$	Miyaji 2000	2×10^{-15}	3×10^{-14}	5×10^{-8}	0.64
$(\log N_H)^8$	Miyaji 2000	3×10^{-17}	3×10^{-16}	5×10^{-8}	0.56
GSH01A	Miyaji 2000	4×10^{-10}	5×10^{-15}	9×10^{-9}	0.44
U03	Miyaji 2000	3×10^{-11}	3×10^{-10}	3×10^{-6}	0.84

survey (e.g. the XMM-LSS, or the XBS Della Ceca et al. 2004) would provide a much better constraint on the AGN population in this flux regime. In the 0.2–0.5 keV band, the Ueda et al. (2003) XLF model reproduces the source counts well in the $5 \times 10^{-16} < S_{0.2-0.5} < 2 \times 10^{-14}$ erg s^{-1} cm^{-2}flux range. At brighter fluxes there is an apparent surfeit sources in the XMM-DODS sample. The four N_H models under-predict the number of these bright, soft X-ray detected sources by a factor of 3–5 at $S_{0.2-0.5} \sim 5 \times 10^{-14}$ erg s^{-1} cm^{-2}. At the faintest 0.2–0.5 keV fluxes there is also an excess in the sample compared to the models, although this excess is less prominent.

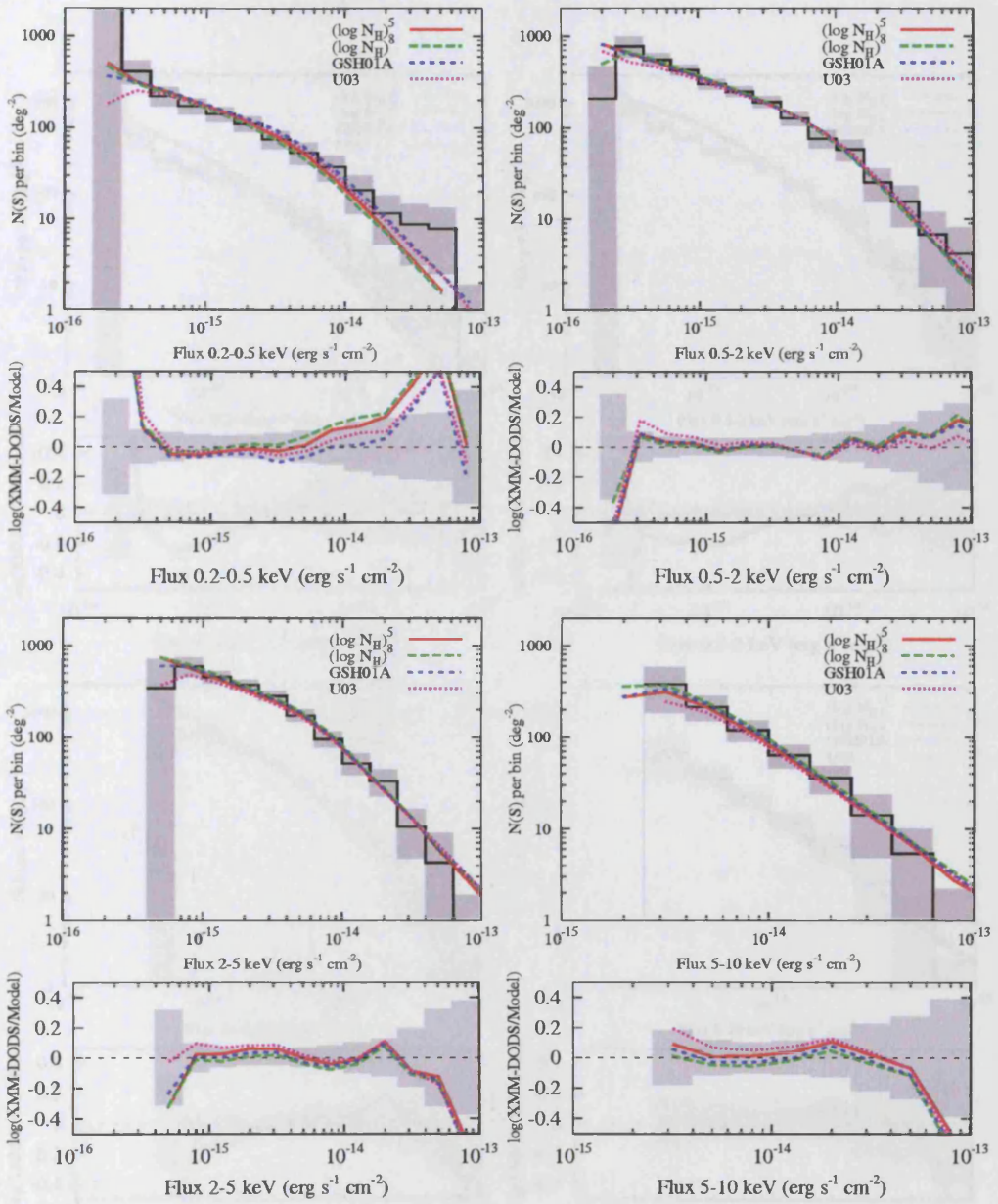

Figure 5.9: Differential source counts in the XMM-DODS sample in comparison to the predictions of several population models. Results are shown for the four energy bands: 0.2–0.5 (top left), 0.5–2.0 (top right), 2.0–5.0 (bottom left) and 5–10 keV (bottom right). The histograms (with shaded $\sqrt{N}$ uncertainty estimates) are for the single-band selected sourcelists in the whole XMM-DODS sample. The curves show the predictions of the four N_H models combined with the Ueda et al. (2003) XLF model. Note that the population models have been adjusted to match the *integral* source counts (in the multi-band selected XMM-DODS sample) at a 0.5–2 keV band flux of 2×10^{-15} erg s^{-1} cm^{-2}.

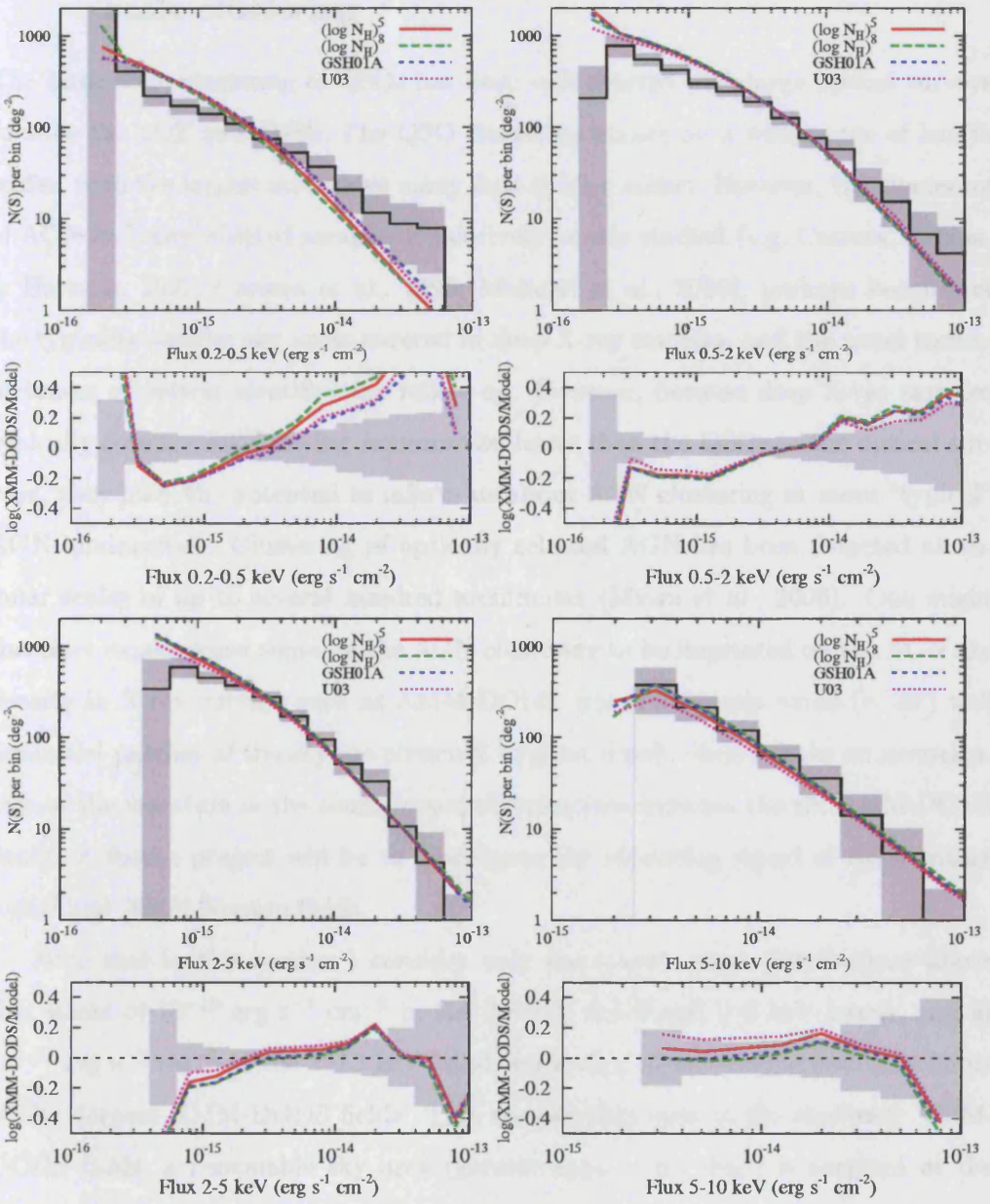

Figure 5.10: Same as figure 5.9, but for model populations generated according to the Miyaji et al. (2000) LDDE1 XLF model.

5.5.4 Field to field density variations as an indicator of large scale clustering

The large scale clustering of QSOs has been well charted with large optical surveys such as the 2QZ and SDSS. The QSO clustering occurs on a wide range of length scales, with the largest structures many tens of Mpc across. However, the clustering of AGN in X-ray selected samples is relatively poorly studied (e.g. Carrera, Fabian, & Barcons, 1997; Carrera et al., 1998; Mainieri et al., 2005), perhaps because of the typically smaller sky areas covered in deep X-ray samples, and the usual incompleteness of optical identification follow up. However, because deep X-ray samples typically contain AGN having luminosities lower than the QSOs in the optical surveys, they have the potential to inform us about AGN clustering at more "typical" AGN luminosities. Clustering of optically selected AGN has been detected at angular scales of up to several hundred arcminutes (Myers et al., 2006). One might therefore expect some signal of the AGN clustering to be imprinted on the AGN sky density in X-ray surveys such as XMM-DODS, where relatively small ($\sim 30'$) well separated patches of the sky are observed to great depth. Here I make an investigation of the variation in the source count distributions between the six XMM-DODS fields. A future project will be to investigate the clustering signal of AGN within individual *XMM-Newton* fields.

Note that in this section I consider only the source count distributions above flux limits of 10^{-15} erg s^{-1} cm^{-2} in the 0.2–0.5, 0.5–2 and 2–5 keV bands, and at 10^{-14} erg s^{-1} cm^{-2} in the 2–10 keV band, somewhat above the nominal flux limits of the deepest XMM-DODS fields. This ensures that even in the shallower XMM-DODS fields, a reasonable sky area (greater than ~ 0.1 deg^2) is sampled at the faintest fluxes (see figure 5.7).

First, I compare the integral number of sources above several flux limits in each XMM-DODS field and in each of the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands. The six XMM-DODS fields cover different sky areas, and have a range of exposure

times. It is vital to account for these factors to allow a direct comparison the source counts from field-to-field. The weighted contribution to the integral source counts of each source is taken to be the inverse of the sky area (in deg^2) in which it could have been detected. The weight for each XMM-DODS source in each energy band is calculated from the sky area covered at the sources's flux in the field where the source is found (see section 5.5.2). The integral source counts in each field above a certain flux are then computed by summing the weights of all sources that are detected above that flux.

Figure 5.11 shows the integral source counts in each of the XMM-DODS fields above several limiting fluxes for the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands. It can be seen in the centre panels of figure 5.11 that there is a considerable amount of scatter of the source counts in each field about the mean value. However, even if the X-ray sources were evenly distributed over the sky one would still expect a certain amount of scatter between fields purely from Poissonian counting uncertainties. I have calculated the likelihood that such statistical uncertainties can account for the scatter measured between the XMM-DODS fields. To do this, I have measured the degree of scatter seen in the output simulated source populations, in which the positions of the input sources are completely random.

I define the statistic $\Delta(> S)$ as a simple numerical measure of the field-to-field scatter in the numbers of sources above a given flux.

$$\Delta(> S) = \sum_{i=1}^{6} \frac{N_i(> S) - N_{mean}(> S)}{\sigma_{Ni}(> S)} \tag{5.7}$$

where $N_i(> S)$ is the integral source counts above flux S in the i^{th} field, $\sigma_{Ni}(> S)$ is the counting uncertainty in the latter value, and $N_{mean}(> S)$ is the integral source counts above flux S averaged over all the XMM-DODS fields. I compute $\Delta(> S)$ for the XMM-DODS sample at several limiting fluxes in each energy band. $\Delta(> S)$ is then computed for each of the 800 (= 2 XLF models $\times$ 4 N_H models $\times$ 100 iterations) simulated AGN populations. The fraction of the simulated populations

with scattering statistic $\Delta(> S)$ greater than some value is equivalent to the probability of finding that degree of apparent scatter between the fields of a completely randomly distributed population. Therefore, the confidence that one can reject the null hypothesis (that the scatter between the six XMM-DODS fields is caused by counting uncertainties rather than cosmic variance) is equal to the fraction of simulated populations with $\Delta(> S)$ lower than the value measured in the XMM-DODS sample.

The lower panels of figure 5.11 show the results of this comparison in each of the 0.2–0.5, 0.5–2, 2–5, and 5–10 keV energy bands. There is strong evidence for a significant (null hypothesis rejected with $> 99\%$ confidence) field-to-field variation in source density between the six XMM-DODS fields in the 0.2–0.5, 0.5–2 and 2–5 keV energy bands. The peak field-to-field scatter occurs at faint fluxes ($\geq 2 - 510^{-15}$ erg s^{-1} cm^{-2}) in the 0.5–2 and 2–5 keV bands, where the hypothesis that the scatter can be due to Poissonian noise is rejected with 99.8% confidence.

There is a trend for the excess scattering of the integral source counts to decrease toward high fluxes. This effect could be a sign of a true dependence of the clustering signal on flux. Alternatively, it could be due to the lower source counts (and therefore poorer counting statistics) at higher fluxes. In order to investigate this further, I have examined the scatter of the *differential* source counts between the six XMM-DODS fields. Figure 5.12 shows the differential source counts in each XMM-DODS field. The plots have been contsructed in a similar way to those in figure 5.11. There are fewer objects per bin than in the integral case, meaning that the Poissonian uncertainties are larger. Therefore this test is sensitive to only rather large fluctuations in the source counts. However, this comparison does highlight which flux bins are contributing to the differences seen in figure 5.11. For example, it can be seen that in the 2–5 keV band the CDFN field is under-dense compared to the XMM-DODS average at all fluxes. In contrast, the LH field has a larger than average number of sources at most flux levels, and in each energy band.

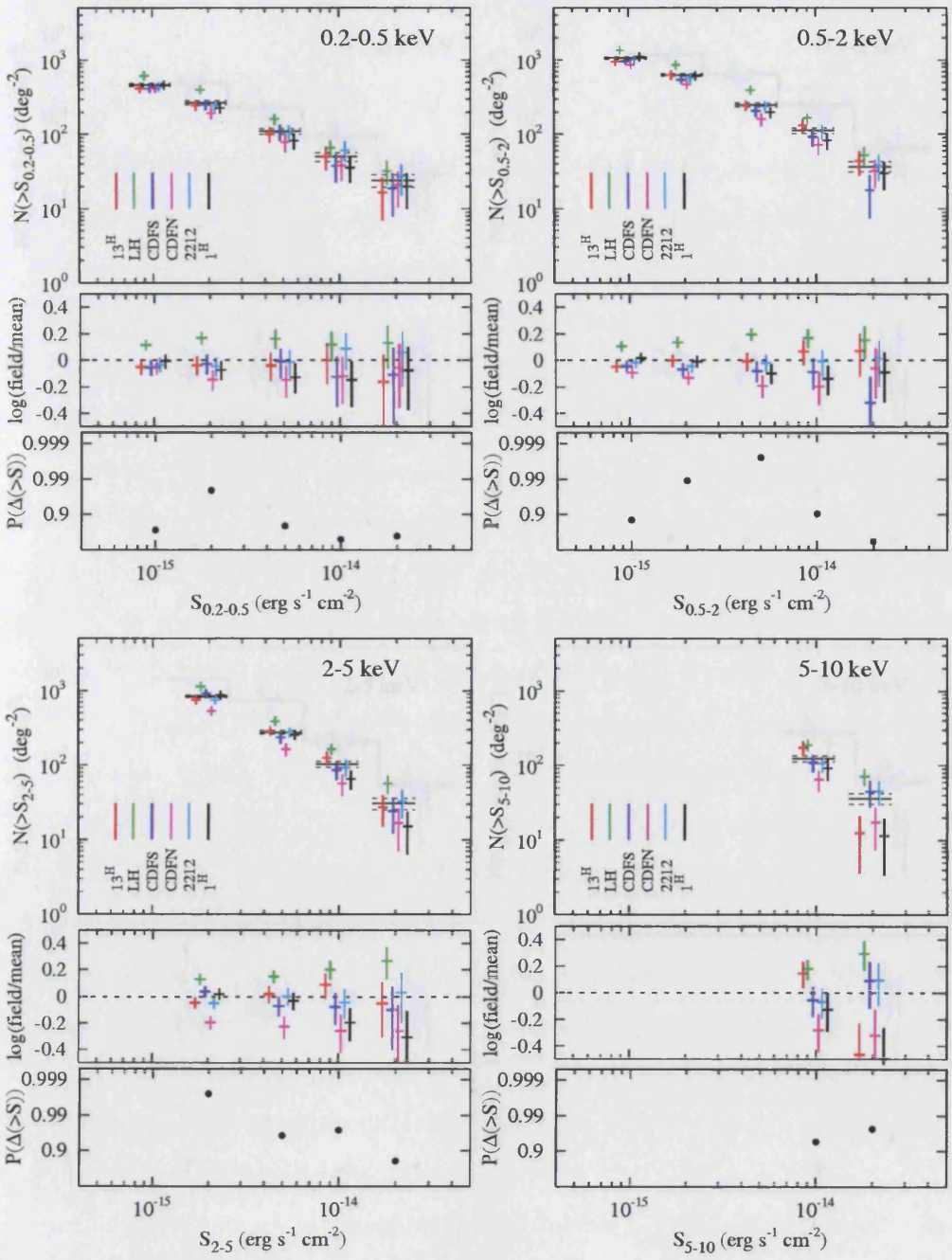

Figure 5.11: The field-to-field scatter in the integral source counts, above a number of flux limits. Each set of three panels shows the following information for a single energy band. *Upper panel:* Vertical bars (offset horizontally for clarity) show the integral source counts in each XMM-DODS field, and the corresponding $\sqrt{N}$ uncertainties. Horizontal bars show the integral source counts in the total XMM-DODS sample (with uncertainties). *Middle panel:* Ratio of the integral source counts in each field to the mean. *Lower panel:* Confidence level $P(\Delta(>S))$ that the field-to-field scatter in the integral source counts is larger than that expected from just Poissonian uncertainty.

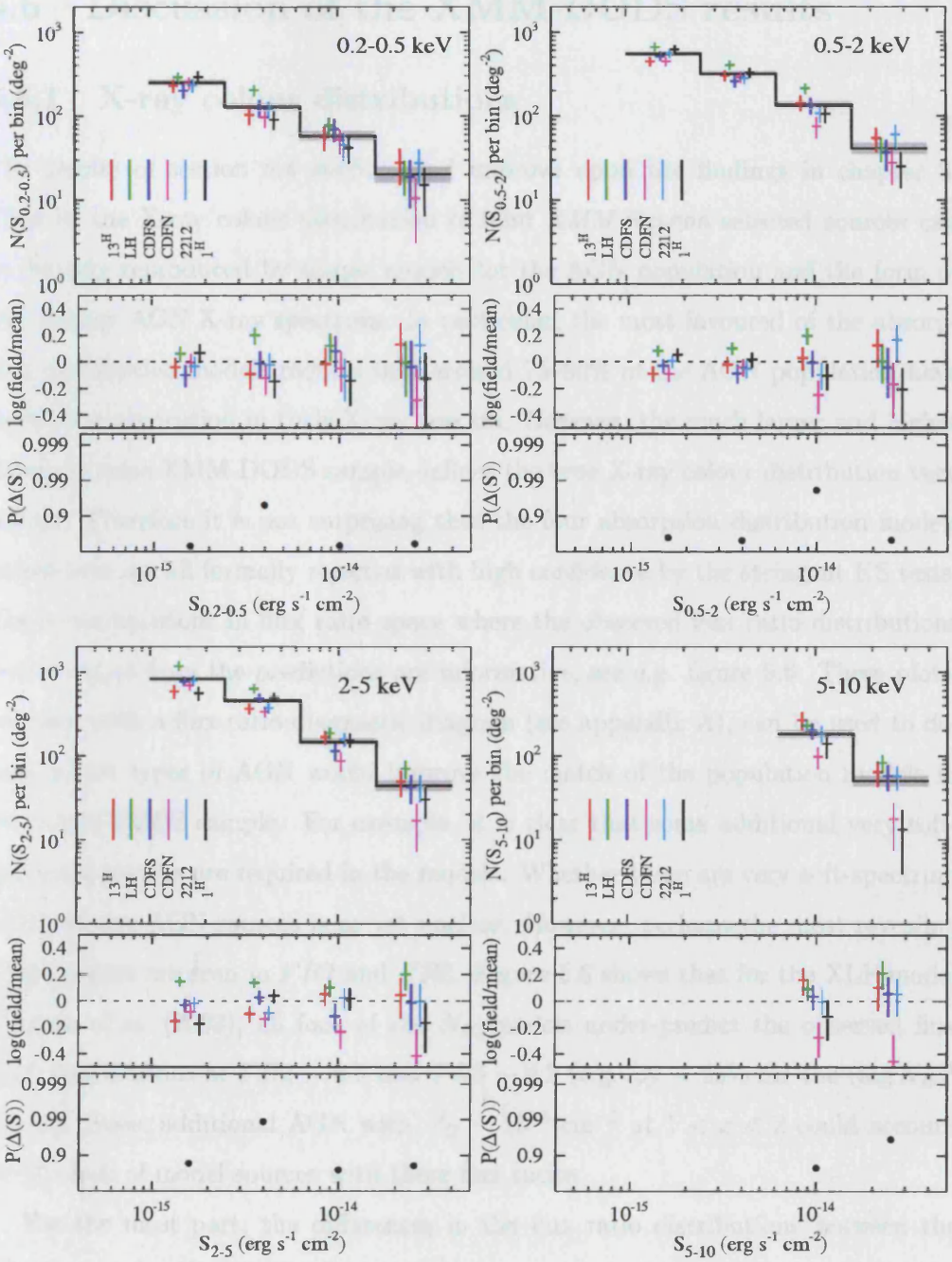

Figure 5.12: The field-to-field scatter in the binned differential source counts. Each set of three panels shows the following information for a single energy band. *Upper panels:* Vertical bars (offset horizontally for clarity) show the differential source counts in each XMM-DODS field, and the corresponding $\sqrt{N}$ uncertainties. The histogram shows the differential source counts in the total XMM-DODS sample with shaded uncertainties. *Centre panels:* Ratio of the differential source counts in each field to the total. *Lower panels:* Confidence level $P(\Delta(S))$ that the field-to-field scatter in the differential source counts is larger than that expected from just Poissonian uncertainty.

5.6 Discussion of the XMM-DODS results

5.6.1 X-ray colour distributions

The results of section 5.4 confirm and improve upon the findings in chapter 3. That is, the X-ray colour distribution of faint *XMM-Newton* selected sources can be broadly reproduced by simple models for the AGN population and the form of the average AGN X-ray spectrum. In particular, the most favoured of the absorption distribution models require that around 75–80% of the AGN population have significant absorption in their X-ray spectra. However, the much larger and higher signal-to-noise XMM-DODS sample defines the true X-ray colour distribution very tightly. Therefore it is not surprising that the four absorption distribution models tested here are all formally rejected with high confidence by the stringent KS tests. The exact locations in flux ratio space where the observed flux ratio distributions deviate most from the predictions are informative, see e.g. figure 5.6. These plots, together with a flux-ratio diagnostic diagram (see appendix A), can be used to deduce which types of AGN would improve the match of the population models to the XMM-DODS sample. For example, it is clear that some additional very soft-spectrum sources are required in the models. Whether these are very soft-spectrum AGN, or non-AGN sources is as yet unclear. However, perhaps the most revealing discrepancies are seen in $FR2$ and $FR3$. Figure 5.6 shows that for the XLF model of Ueda et al. (2003), all four of the N_H models under-predict the observed flux ratio distributions at $FR2 \sim 0.5$ and $FR3 \sim 0.2$ (e.g. by $\sim 25\%$ for the $(\log N_H)^8$ model). Some additional AGN with $N_H \sim 10^{23}$ cm^{-2} at $1 < z < 2$ could account for the lack of model sources with these flux ratios.

For the most part, the differences in the flux ratio distributions between the models and the XMM-DODS sample make up only a fairly small fraction of the population. The differences could be explained by a combination of AGN with more complex spectra, as well as a small number of non-AGN sources contaminating the XMM-DODS sample. The best matching N_H distribution, the $(\log N_H)^8$ model, does

broadly reproduce the pattern of X-ray colours observed in the AGN population.

Only a small fraction of the XMM-DODS sample have redshift determinations so I was not able to expand on the analysis of chapter 4. This will have to wait several years until a higher fraction of the XMM-DODS sample are optically identified. However, the two N_H models which were most favoured by the results of chapter 4 (the $(\log N_H)^8$ and *GSH01A* models), are also preferred by the X-ray colour analysis in this chapter.

5.6.2 Source count distributions

In section 5.5, I showed that the source counts predicted by model populations following the LDDE1 XLF of Miyaji et al. (2000) provided a poor match to the source counts in the XMM-DODS sample. In contrast the AGN populations following the Ueda et al. (2003) XLF model do provide a reasonable description of the source counts in the XMM-DODS sample, at least at fluxes fainter than 10^{-14} erg s^{-1} cm^{-2}. At high fluxes, especially in the 0.2–0.5 keV band, some modification of the AGN population models may be required in order to reproduce fully the observed source counts, where there is an excess of very bright sources. A wider area *XMM-Newton* sample is required in order to fully test this. The excess of bright sources in the 0.2–0.5 keV band may be caused by AGN with very soft spectra. These soft spectrum objects are not represented in my model AGN populations. Alternatively, they could be non-AGN sources, such as unresolved groups/clusters of galaxies, which because of incomplete identification, remain in the XMM-DODS AGN sample. It is evident that the differences between the four N_H models have a far smaller effect on the source count predictions than the differences between the two XLF models that I have tested here. This is especially true in the hardest X-ray band where for example, the relative number of sources with intermediate absorption (which is where the N_H models tested here differ most) is unimportant. However, at softer energies some discrimination is possible between the models; the *U03* N_H model

provides a somewhat poorer description of the XMM-DODS source counts than the other three N_H models tested here.

5.6.3 Field to field source density variations

In section 5.5.4 I showed that there is strong evidence for source density differences between the six XMM-DODS fields that cannot be explained by Poissonian fluctuations alone. The CDFN field is significantly under-dense, and the LH over-dense relative to the mean source density over all six XMM-DODS fields. A future project will be to examine the angular clustering *within* the XMM-DODS fields, which will probe AGN clustering at somewhat smaller length scales.

Chapter 6

Discussion and Conclusions

6.1 Tracing the absorbed AGN population

The findings of this thesis are in broad agreement with XRB synthesis models dating from before the *XMM-Newton/Chandra* era (e.g. Comastri et al., 1995; Gilli et al., 2001), i.e. absorbed AGN have a similar distribution in redshift/luminosity space as their unabsorbed counterparts, and that absorbed AGN are $\sim 3-4$ times as numerous as unabsorbed AGN. In particular, my results imply that contrary to some recent *Chandra* results (e.g. Ueda et al., 2003; Barger et al., 2005; Treister & Urry, 2005), there *does* exist a large population of heavily absorbed luminous AGN at high redshifts. I have illustrated how vital it is to take account of the complex selection functions in X-ray selected AGN samples. It is also clear from my X-ray colour analyses that the AGN population contains objects with a broad range of absorbing column densities, including many AGN with intermediate absorption. The X-ray properties of the vast majority of the AGN can be well described by a simple spectral model; an absorbed or unabsorbed power law with a component of reflected emission.

6.2 The luminous absorbed AGN population

Rather few luminous absorbed (type-2) QSOs have been found in X-ray surveys to date. There are certainly fewer than predicted by XRB synthesis models in which a large fraction of the XRB is made up of luminous but highly absorbed AGN at redshifts 1.5–2.5 (Comastri et al., 1995; Gilli et al., 2001). In response to the lack of type-2 QSOs, models have been devised in which the fraction of AGN with significant absorption declines with increasing luminosity (Barger et al., 2005; Ueda et al., 2003). The physical interpretation is that highly luminous QSOs have the power to remove a significant fraction of the surrounding material, effectively increasing the opening angle of any circumnuclear structure, a so called "receding torus" model (e.g. Lawrence, 1991; Simpson, Rawlings and Lacy, 1998; Simpson, 2005). The X-ray samples detected in *Chandra* pencil-beam studies *do* contain large numbers of absorbed AGN, but they lie at lower redshifts, and have lower luminosities than the peak in type-1 QSO activity ($L_X \sim 10^{44}$ erg s^{-1}, $z \sim 1.5-2$). It is this low redshift, low luminosity absorbed population which is postulated to take the place of the type-2 QSOs in making up the hard spectrum of the XRB. These findings, if they are taken at face value, raise important questions about the cosmic history of the AGN population. The implication is that, with soft X-ray and optical surveys, we have already detected the majority of intrinsically luminous accretion powered objects in the form of type-1 QSOs. In such a scheme, the total energy output that is powered by accretion, integrated over cosmic timescales, is significantly reduced from the predictions of the simplest "unified" schemes.

A few rare examples of type-2 QSOs, selected in X-ray and/or optical surveys have been reported in the last few years (e.g. Norman et al., 2002; Della Ceca et al., 2003; Loaring, Page & Ramsay, 2003; Gandhi et al., 2004; Ptak et al., 2006). In fact, one of these, the Norman et al. (2002) object, appears in the *XMM-Newton* observations of the CDFS(see chapter 4). The conclusion many have drawn from this is that type-2 QSOs are less numerous than the large population predicted by say

the Gilli et al. (2001) XRB synthesis model. However, recent mid-infrared surveys with *Spitzer* have started to reveal the existence of significant numbers of type-2 QSOs. Surveys in the mid-infrared are sensitive to emission originating in dusty material that is heated by a compact heat source, i.e. a powerful AGN. Martinez-Sansigre et al. (2005) have demonstrated that type-2 QSOs at $z > 2$ can be selected efficiently by choosing objects bright at 24μm, but faint at near-infrared and radio wavelengths. Using these criteria, they identified a population of luminous obscured QSOs between 1 and 3 times as numerous as the population of luminous type-1 QSOs found by optical surveys (e.g. Croom et al., 2004). The absorbed fraction of 50–75% found by Martinez-Sansigre et al. (2005) is comparable to the absorbed fraction at high luminosities of $\sim 75\%$ predicted by the two N_H models (namely the $(\log N_H)^8$ and *GSH01A* N_H models) which best match the pattern of absorption in the X-ray samples I have examined in this study.

6.3 Obscured black hole growth over cosmic timescales

I have shown that the source counts, and X-ray colour distribution of AGN in the *XMM-Newton* samples are best matched by population models in which obscured AGN are ~ 3 times as populous as unobscured AGN at all redshifts and luminosities. This result implies that most ($\sim 75\%$) of supermassive black hole growth is obscured. How does this compare to other measures of the accretion history of the Universe? Various studies have attempted to reconcile the observed accretion powered luminosity density at high redshifts, with the locally observed relic black hole mass function (e.g. Fabian & Iwasawa 1999; Yu & Tremaine 2002; Elvis, Risaliti & Zamorani 2002; Marconi et al. 2004). Yu & Tremaine (2002) found that even with a high accretion efficiency (~ 0.1), the luminosity density of quasars seen in optical surveys was sufficient to have produced the locally measured black hole mass

density. This result left little room for any additional obscured accretion (and hence black hole growth) as required by XRB synthesis models (e.g. Gilli et al., 2001). However, such studies are dependent on two crucial assumptions; the correction factor used to convert optical luminosity to bolometric luminosity and the local black hole mass density. More recent studies have revisited this question with a revised set of assumptions. For example, the study of Marconi et al. (2004) uses a more detailed bolometric correction method, and a higher estimate of the local black hole mass density than that calculated by Yu & Tremaine (2002). Marconi et al. (2004) then found that the AGN population model of Ueda et al. (2003, XLF and N_H function) was consistent with the local black hole mass function if the mean accretion efficiency is ~ 0.08. However, as I have shown in the previous chapter, the observed numbers of luminous, absorbed AGN are substantially higher than the predictions of the *U03* N_H model. Coupled with the Martinez-Sansigre et al. (2005) result, the implication is that the mean accretion efficiency of AGN is higher than the Marconi et al. (2004) result, or alternatively that the local black hole mass density has been underestimated.

The tight correlation observed between supermassive black hole mass and the velocity dispersion (and hence mass) in host galaxy spheroids (Magorrian et al., 1998; Merritt & Ferrarese, 2001), implies a close connection between the growth of central black holes and the star formation which builds their host spheroids (e.g. Silk & Rees, 1998). Current theories of galaxy formation require some "feedback" from AGN that is able to regulate the star formation in their host galaxies. Radio jets or AGN winds/outflows are typically suggested as candidate physical processes by which accretion power from an AGN can influence its galactic environment. Recent galaxy formation models contain AGN feedback terms in order to model the growth and evolution of galaxies over cosmic history (e.g. Springel, Di Matteo & Hernquist, 2005). Clearly a good knowledge of the *total* AGN population, not just the easily detected "type-1" unabsorbed objects, is required for the AGN feedback to be modelled properly.

6.4 Summary

I have compared the X-ray properties of a large sample of AGN with the predictions of a number of model AGN populations, utilising Monte Carlo simulations to account for the selection function and particulars of the *XMM-Newton* observations. I find no evidence for a decline in the fraction of AGN with significant absorption at high luminosities that has been reported by many authors. The N_H distribution models which most closely match the pattern of absorption in the AGN samples I have studied are independent of both redshift and luminosity over the effective ranges probed by the observations ($0 < z < 4$, $10^{40} < L_{2-10} < 10^{45}$ erg s^{-1}).

The XMM-CDFS sample contains at least 23 heavily absorbed AGN with QSO-like intrinsic X-ray luminosities. I postulate that nearly half of the objects without redshift determinations in the XMM-CDFS are also absorbed QSOs, and the reason they are without photometric-redshift estimates is because they lie in the "photo-z desert" ($1.5 < z < 2.5$). Spectroscopic observations of these sources will be required to confirm this hypothesis.

6.5 Future directions and outlook

The XMM-CDFS and XMM-DODS samples can be further exploited in order to understand better the nature of the sources which make up the bulk of the XRB.

In this thesis I have compared the relative ability of several N_H models to reproduce the distribution found in the faint X-ray population, but have not yet determined a single N_H distribution which provides the best fit to the observations. A possible approach to do this is by starting with one of the most successful model N_H distributions, and then iteratively modifying the distribution so that it better matches the X-ray colour distribution seen in the XMM-DODS sample (for example by minimising the χ^2 between the binned flux ratio distributions of the real and model populations). As suggested by my examiners, the computing time required

for such an iterative fitting process could be reduced by "recycling" sources from the library of simulated populations.

Except in the very local Universe, the size of the Compton thick AGN population is only very poorly determined. Being able to constrain the number of Compton-thick objects at higher redshifts is very important if we are to account for all the sources which make up the XRB. One would expect a survey the size and depth of XMM-DODS to contain a number of Compton-thick AGN, either detected by their attenuated direct emission, and/or by scattered X-ray emission. Indeed, in the XMM-CDFS sample there are at least 10 AGN with columns $\geq 10^{24}$ cm^{-2}. A further project will thus be to select a sample of candidate Compton-thick AGN from the XMM-DODS sample on the basis of their X-ray colours and optical faintness. These objects can then be examined in more detail, with a full X-ray spectral analysis and/or with multi-wavelength follow up.

In a similar way, X-ray colour selection criteria can be used to define a subset of XMM-DODS sources with unusual spectra which merit a more detailed examination. For example, there are a number of XMM-DODS sources which have more flux in the 0.2–0.5 keV band than expected given their X-ray colours above 0.5 keV and assuming a simple absorbed power law spectral form. A more detailed spectral analysis should reveal whether the soft excess is due to reprocessed X-rays or an ionised absorber, or alternatively to X-ray emission powered by rapid star formation. The latter can tell us about simultaneous black hole growth and galaxy formation, and the former two possibilities can potentially tell us something about the geometry and composition of AGN tori.

The large FOV and unprecedented throughput of *XMM-Newton* mean that it is accumulating X-ray photons several times faster than *Chandra*. The *XMM-Newton* archive is thus becoming an increasingly attractive resource with which to study the X-ray source population. For instance, AGN samples from the deepest *XMM-Newton* surveys (i.e. XMM-DODS) can be combined with samples found in shallower but wider area surveys (e.g. the $\sim$1 deg^2 Subaru-XMM deep field, the $\sim$2 deg^2

XMM-COSMOS field, and the ~8 deg^2 XMM-LSS field). These shallow but wide surveys can improve the sampling of the redshift/luminosity plane, in particular by detecting examples of the extremely luminous QSO population ($L_X > 10^{45}$ erg s^{-1}), which have a rather low sky density. It will be very interesting to see whether the luminosity-independent N_H distribution models which match the AGN population below 10^{45} erg s^{-1} can also reproduce the absorption distribution in the AGN at extreme luminosities. However, deep optical/NIR imaging and spectroscopic follow-up programs for these wide areas are expensive in telescope time, and completely identified X-ray samples are not yet available. Therefore in the short term, X-ray colour analyses similar to that used in chapter 5 may be the best way to exploit these wider area *XMM-Newton* data-sets.

Appendix A

X-ray colour-colour diagnostic plots

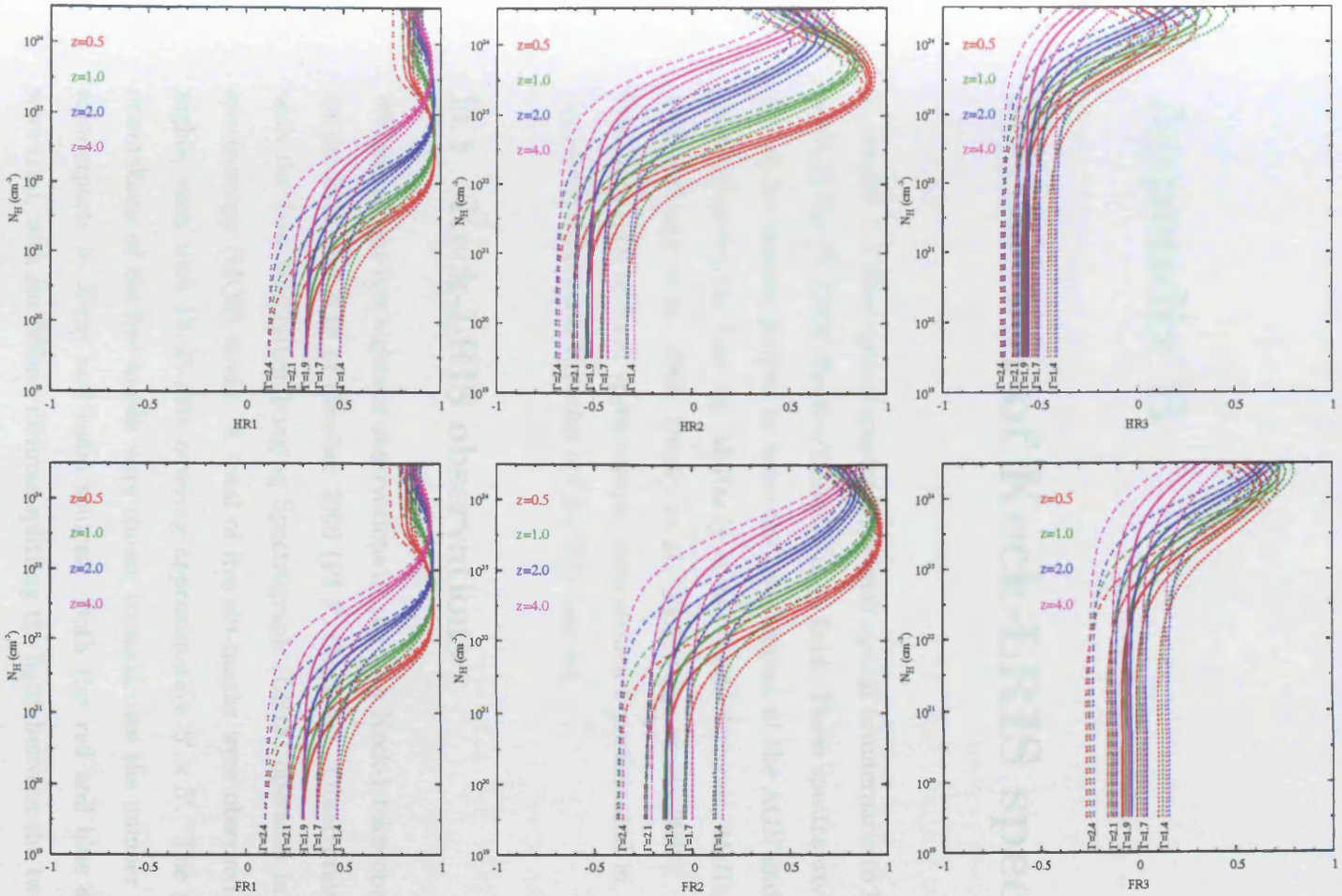

Figure A.1: The dependence of hardness ratios (top row) and flux ratios (bottom row) on the AGN spectral parameters N_H, z and Γ. The assumed spectral model is the "absorbed power law plus reflection" spectral model (see section 3.4.4). The HRs are calculated for a Galactic column of 6×10^{19} cm^{-2}. The four different dash patterns are for intrinsic spectral slopes of 1.4, 1.7, 1.9, and 2.2. The four different line colours are for redshifts of 0.5, 1.0, 2.0, and 4.0.

Appendix B

Reduction of Keck-LRIS spectra

In chapter 1, I show optical spectra for several optical counterparts to X-ray selected AGN in the 1^{H} *XMM-Newton/Chandra* deep field. These spectra were collected as part of an ongoing project to trace the co-evolution of the AGN and star forming galaxy populations (see e.g. McHardy et al., 2003; Seymour, McHardy & Gunn, 2004; Loaring et al., 2005; Dwelly et al., 2005; Page et al., 2006). I here give a brief summary of Keck spectroscopic observations I participated in, and describe the reduction process I carried out for this data set.

B.1 Keck-LRIS observations

We carried out two nights of observations at the 10m Keck-1 telescope, Mauna Kea on the 29th and 30th of October 2003 (PI F. Cordova). Observations were made with the Low Resolution Imaging Spectrograph (LRIS) operated in multi object spectroscopy (MOS) mode. A total of five slit-masks were observed over the two nights, each with 15–25 slits covering approximately $8' \times 5'$. The positions and orientations of the five masks were chosen to maximize the number of observable counterparts to X-ray and radio sources. Both the red and blue arms of LRIS were used, with the 560nm dichroic splitting the light between the two arms of the instrument. All masks were observed using the "600/4000" grism in the blue, and

the "400/8500" grating in the red. The wavelength coverage with this particular instrument set-up varies according to the position of the object in the mask, but for slits positioned centrally, the 3500–9500Å interval is covered continuously. For unfavourably positioned objects it leads to a small gap in the wavelength coverage between the red and blue arms. We used 1.0″ slits, which, together with the choice of grating/grism, gave a resolution with FWHM $\sim$ 5Å in the blue, and $\sim$ 9Å in the red (measured from the arc frames used for wavelength calibration).

Dome flat frames were taken at the beginning and/or end of the night with the slitmasks and grism/grating in place in order to measure the relative slitlet thoughput function. Multiple flats were taken for each slitmask to improve the number of counts in the blue/NUV. Wavelength calibration frames were taken either immediately before or after observing each mask. The Ne, Ar, Hg, Zn, and Cd arc lamps were used ensuring that accurate wavelength calibration was possible across the entire wavelength range. Spectrophotometric standard stars (Feige 34, BD284211, and G191B2B) were observed with the 1.0″ long slit at the beginning and end of each night.

The observations were made in fairly good weather conditions. The first night was almost completely photometric, but approximately 2 hours of the second night were lost to thick cloud cover. The lost time meant that we observed 4/6 of the planned slitmasks with 4 × 30 minute exposures, one slitmask was observed with two 30 minute exposures, and one slitmask was unobserved.

B.2 Reduction process

The Image Reduction and Analysis Facility (IRAF Tody, 1986) was used to reduce the spectra, notably the NOAO *twodspec apextract* package, and the LRIS-specific package *wmkolris*. The document "User's Guide to Reducing Slit Spectra with IRAF" Massey, Valdes,& Barnes, was particularly informative.

The red arm of LRIS has a single 2k×2k CCD which is read out through two

amplifiers. The "new" blue camera consists of two 2k×4k CCDs each of which is readout through two amplifiers, and so has four 1k×4k sections with somewhat different bias and gain values. The bias frames show significant structure, and tso the foillowing procedures were used to remove the bias from the science frames. For the blue, a composite surface was fitted to a master bias frame (taking account of the four CCD+amplifier sections), and then this fitted surface was subtracted from the blue science frames. In the red, the science frames were debiased using the *lrisbias* task, which utilises the unexposed overscan sections of the red CCD.

The goal of flatfielding is to determine the relative transmission along each slitlet, as well as to measure the small variations in response from pixel to pixel. At this stage one does not need to determine the throughput of the system with respect to wavelength. The following procedure was carried out for each slitmask, and for both the red and blue arms. Firstly, the (on average 3) flat frames were averaged, with a cosmic ray rejection algorithm. Areas of the averaged flat frame with fewer than 2000 (mean) counts were masked out - below this level, we expect counting nose to dominate the pixel-to-pixel response differences. Areas between the slitlets were also masked out. The slitlets are offset with respect to each other in the dispersion direction, and so each must be treated separately for the flattening. I used the *apnormalise* package to fit a surface to the each flat "stripe", with a high order spline function along the dispersion axis, and a constant along the spatial axis. The measured flat frame was then divided by the fitted surfaces, to give a normalised image (with mean $\sim$ 1). Areas which had ben excluded earlier were then set to 1. Finally, each (previously debiased) science frame was divided by the appropriate normalised flatfield image.

The science frames (four separate frames for four of the masks, and three frames for one slitmask) were averaged, and a cosmic ray rejection algorithm applied. A handful of residual cosmic rays were removed manually using the *imedit* application.

The *apall* application was used to define the source apertures and background regions, fit a low-order trace to the spectral streak, and finally to extract the spectra.

This process was done manually because the faintness of the target sources precludes the use of automatic extraction algorithms. Care was taken to chose source and background apertures that would maximise the signal-to-noise in the extracted 1D spectra. A set of wavelength calibration spectra were extracted from the arc-line frames using the apertures defined for the object spectra. The *identify* package was used to calculate a wavelength solution for each slitlet from the 1D arc-spectra. The appropriate wavelength solutions were then applied to each of the object spectra.

A similar process was carried out to extract spectra for three spectrophotometric standard stars, each of which was observed multiple times. The standard star spectra were then compared to the flux measurements of Oke (1990) in order to calculate the (wavelength dependent) absolute throughput of the telescope+optics+instrument. Telluric absorption features were interpolated over at this stage. The science frames were then flux calibrated using the standard star measurement closest in airmass and time to the observation.

For each of the observed slitmasks, several high signal-to-noise objects having fairly featureless continua spanning the strongest telluric absorption features at 6900Å and 7600Åwere chosen. Care was taken to exclude sources at redshifts where commonly seen emission lines would be expected to fall in these bands. For each slitmask, the spectra of the chosen sources were averaged and the resultant continuum fitted with a low order polynomial. The ratio of the data to the fitted function in the telluric bands is a good estimate of the strength of the atomspheric absorption. The object spectra were divided (only in the telluric absorption bands) by the measured ratio. The advantage of this method is that the correction factors are derived from spectra which have been observed in exactly the same atmospheric conditions as the spetra that are to be corrected.

Appendix C

An optically faint, X-ray bright source in the XMM-CDFS

As detailed in section 4.3.3, there are 50 X-ray sources in the XMM-CDFS sample which remain unidentified. The X-ray brightest of the unidentified objects in the XMM-CDFS sample (hereafter XMM-CDFS-U1) is particularly noteworthy. Figure C.1 shows "postage stamp" images of XMM-CDFS-U1 from X-rays to the MIR. XMM-CDFS-U1 has a 5–10 keV flux of $2.3 \pm 0.2 \times 10^{-14}$ erg s^{-1} cm^{-2} making it the fifth brightest object in the XMM-CDFS sample in this band. XMM-CDFS-U1 is also detected as E-CDFS-437 (Lehmer et al., 2005), which constrains its X-ray position to within 0.6″of $03^{H}32^{m}42.63^{s} - 27°38'16.1''$. There is no object in the COMBO-17 catalogue at this location. However, there is a bright ($H = 19.42 \pm 0.07$) H-band object consistent with this location (Object 495 of Moy et al. 2003). The EIS optical catalogue (Arnouts et al., 2001) also contains a source at this location with $R_{AB} = 24.5 \pm 0.17$, although one would expect a source of that R magnitude to have appeared in the COMBO-17 catalogue. However, a point-like optical counterpart is clearly seen in the "z" ($F850LP$) filter of the GEMS-ACS imaging with $z(AB) = 23.9$ (1″ aperture). There is also a marginal detection at this location in the GEMS-ACS "V" ($F606W$) image, with $V(AB) \sim 27.8$.

This object is a prime candidate for being a type-2 QSO. The extreme X-ray to optical ratio of this object, together with its hard X-ray spectrum, requires that the AGN is very heavily obscured. If this object has similar optical properties to other reported high-z type-2 QSOs, i.e. no blue continuum, but strong narrow emission lines (such as Ly-α, CIV, CIII), then it may be that the ACS V band detection is the integrated light from several narrow emission lines, and would make optical spectroscopic identification of this object feasible. A proposal has been submitted to identify this object with the VLT. If XMM-CDFS-U1 is found to lie at $z > 2$, then it has an intrinsic X-ray luminosity $L_{2-10} > 10^{45}$ erg s^{-1}, similar to the powerful QSO-2 reported by Severgnini et al. (2006).

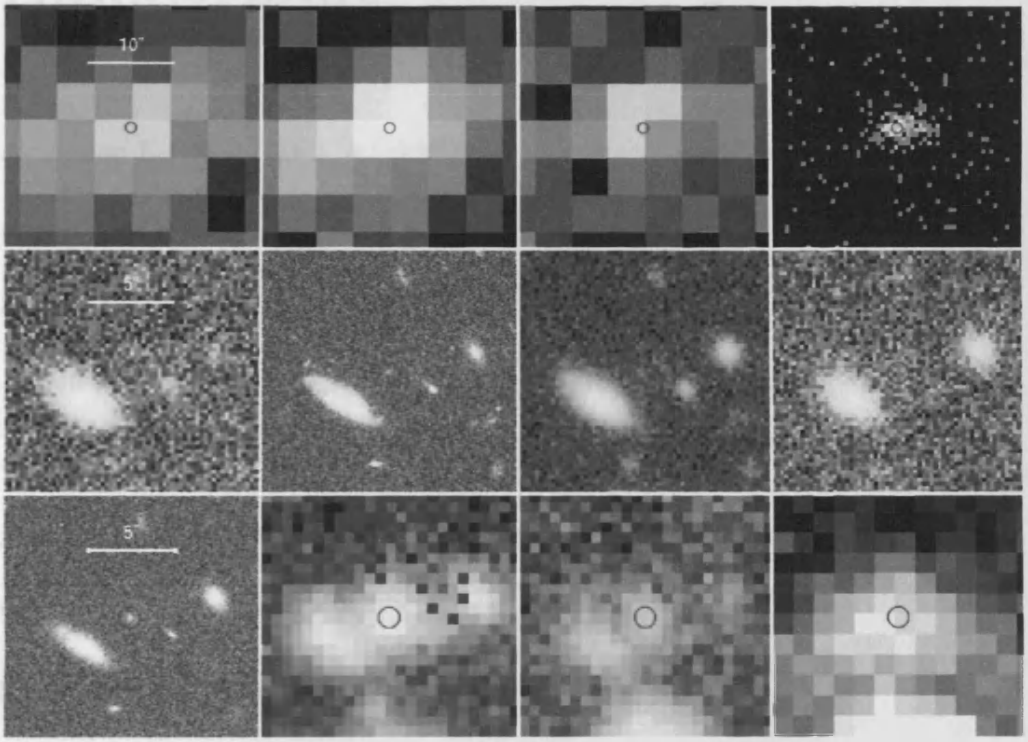

Figure C.1: Multi-wavelength cutout images of the QSO-2 candidate source (E-CDFS-437). Panels are as follows: Top row) *XMM-Newton* EPIC 0.5–2 keV, 2–5 keV and 5–10 keV, E-CDFS 0.5–10 keV. Middle row) WFI-B, HST-ACS-F550W, WFI-R, WFI-I. Bottom row) HST-ACS-F850LP, IRAC-4.5μm,IRAC-8.0μm, MIPS-24μm. The circle in each image shows the 0.6″ E-CDFS-437 position error. The *Spitzer* images are taken from the legacy data products archive (Dickinson et al., 2006).

Bibliography

Abazajian K., et al., 2003, AJ, 126, 2081

Alexander D.M., et al., 2003, AJ, 126, 539

Alonso-Herrero A., et al., 2006, ApJ, 640, 167

Antonucci R.R.J., Miller, J.S., 1985, ApJ, 297, 621

Antonucci R.R.J., 1993, Annu.Rev. A&A, 31, 473

Arnouts S., et al., 2001, A&A, 379, 740

Balucinska-Church M., McCammon D., 1992, ApJ, 400, 699

Barger A., et al. 2003, AJ, 126, 632

Barger A., et al., 2005, AJ, 129, 578

Bauer, F.E., et al., 2004, AJ, 128, 2048

Bautz M., et al. 1998, Proc. SPIE, 3444, 210

Becker R.H., White R.L., Helfand D.J., 1995, ApJ, 450, 559

Blandford R.D., McKee C.F., 1982, ApJ, 255, 419

Boyle B.J., Shanks T., Peterson B.A., 1988, MNRAS, 235, 935

Boyle B.J., Griffiths R.E., et al. 1993, MNRAS, 260, 49

Boyle B.J., Shanks T., et al. 1994, MNRAS, 271, 639

Braito V., et al., 2005, ApJ, 621, L97

Branduardi-Raymont et al. 1994, MNRAS, 270, 947

Caccianiga A., et al. 2004, A&A, 416, 901

Caldwell J.A.R, et al., preprint, astro-ph/0510782

Carrera F.J., Fabian A.C., Barcons X., 1997, MNRAS, 285, 820

Carrera F.J., et al., 1998, MNRAS, 299, 229

Chiappetti et al., 2005, A&A 439, 413

Ciliegi P., et al., 1995, MNRAS, 277, 1463

Civano F., Comastri A., Brusa M., 2005, MNRAS, 358, 693

Clavel J., Schartel N., Tomas L., 2006, A&A, 446, 439

Comastri A., et al., 1995, A&A, 296, 1

Comastri A., et al., 2002, ApJ, 571, 771

Cowie L.L., et al., 2003, ApJ, 584, L57

Croom S.M., et al., 2002, MNRAS, 335, 459

Croom S.M., et al., 2004, MNRAS, 349, 1397

den Herder J.W., et al., 2001, A&A, 365, L7

Della Ceca R., et al., 2003, A&A, 406, 555

Della Ceca R., et al., 2004, A&A, 428, 383

Dickinson M., et al., 2006, see http://data.spitzer.caltech.edu/popular/goods/

Dwelly T., et al., 2005, MNRAS, 360, 1426

Dwelly T., Page M.J, 2006, MNRAS, 372, 1755

Eddington A.S., 1913, MNRAS, 73, 359

Elvis M., et al., 1994, ApJS, 95, 1

Elvis M., 2000, ApJ, 545, 63.

Elvis M., Risaliti G., Zamorani G., 2002, ApJ, 565, L75

Fabian A.C., Iwasawa K., 1999, MNRAS, 303, L34

Fasano G., Franceschini A., 1987, MNRAS, 225, 155

Ferrarese L., Ford H., Jaffe, W., 1996, ApJ, 470, 444

Franceschini A., Braito V., Fadda D., 2002, MNRAS, 335, L51

Gallagher, S.C., et al., 2002, ApJ, 567, 37

Gallagher, S.C., et al., 2006, ApJ, 644, 709

Gandhi, P., Fabian, A., 2003, MNRAS, 339, 1095

Gandhi, P., et al., 2004, MNRAS, 348, 529

Gendreau et al., 1995, PASJ, 47, 5

Georgantopoulos I., et al., 2004, MNRAS, 352, 91

George I.M., Fabian A.C., Nandra K., 1990, MNRAS, 242, 28L

Giacconi R., et al., Physical Review Letters, 1962, 9, 439

Giacconi R., et al., 2002, ApJS, 139, 369

Giavalisco M., et al., 2004, ApJ, 600, L93

Gilli R., Salvati M., Hasinger G., 2001, A&A, 366, 407

Gilli R., et al., 2003, ApJ, 592, 721

Gilli R., 2003, preprint, astro-ph/0303115

Gilli R., et al., 2005, A&A, 430, 811

Granato G.L., Danese L., Franceschini A., 1997, ApJ, 486, 147

Green P.J., Mathur S., 1996, ApJ, 462, 637

Gruber, D.E., Matteson J.L., Peterson L.E., 1999, ApJ, 520, 124

Grupe D., Mathur S., Elvis M., 2003, AJ, 126, 1159

Gunn K.F., et al., in prep.

Hasinger G. et al., 1998, A&A, 329, 482

Hasinger G., 2000, LNP, 548, 423

Hasinger G., Miyaji T., Schmidt M., 2005, A&A, 441, 417

Heckman T.M., et al., 1986, ApJ, 311, 526

Ishisaki Y., et al., 2001, PASJ, 53, 445

Jansen F., et al., 2001, A&A, 365, L1

Jones L.R., et al., 1997, MNRAS, 285, 547

Jones L.R., McHardy I.M., et al., 2002, MNRAS, 334, 219

Kinkhabwala et al., 2002, ApJ, 575, 732

Komossa S., et al., 2006, ApJ, 639, 710

Krolik J.H., et al., 1991, ApJ, 371, 541

Krolik J.H., Madau P., Zycki P.T., 1994, ApJ, 420, L57

La Franca et al., 2005, ApJ, 635, 864

Lamastra A., Perola G.C., Matt, G., 2006, A&A, in press

Lawrence A., 1991, MNRAS, 252, 586

Le Fevre O., et al., 2004, A&A, 428, 1043

Lehmann I., et al., 2001, A&A, 371, 833

Lehmer B.D., et al., 2005, ApJS, 161, 21

Lira P., et al., 2002, MNRAS, 330, 259

Loaring N.S., Page M.J., Ramsay G.T.B., 2003, MNRAS, 345, 865

Loaring N.S., et al., 2005, MNRAS, 362, 1371

Lumb D.H., et al., 2002, A&A, 389, 93

Martini P., et al., 2003, ApJ, 589, 774

McHardy I.M., et al., 1998, MNRAS, 295, 641

McHardy I.M., Gunn, K.F., et al., 2003, MNRAS, 342, 802

Magdziarz P., Zdziarski A.A., 1995, MNRAS, 273, 837

Magorrian, J., et al., 1998, AJ, 115, 2285

Maiolino R., Rieke G.H., 1995, ApJ, 454, 95

Mainieri V., et al., 2002, A&A, 393, 425

Mainieri V., et al., 2005, A&A, 437, 805

Manners J.C., et al., 2003, MNRAS, 343, 293

Marconi A., et al., 2004, MNRAS, 351, 169

Martinez-Sansigre A., et al., 2005, Nature, 436, 666

Mateos S., et al., 2005a, A&A, 433, 855

Mateos S., et al., 2005b, A&A, 444, 79

Merritt D., Ferrarese L., 2001, MNRAS, 320, L30

Miyaji T., Hasinger G., Schmidt M., 2000, A&A, 353, 25

Moy E., et al., 2003, A&A, 403, 493

Myers, A.D., et al., 2006, ApJ, 638, 622

Nandra K., et al., 2005, MNRAS, 356, 568

Norman C., et al., 2002, ApJ, 571, 218

Oke J.B., 1990, AJ, 99, 1621

Osterbrock D.E., 1988, PASP, 100, 412

Page M.J., et al., 1997, MNRAS, 291, 324

Page M.J., Mittaz J.P.D., Carrera F.J., 2001, MNRAS, 325, 575

Page M.J., et al., 2003, AN, 324, 101

Page M.J., et al., 2006, MNRAS, in press.

Perola G.C., et al., 2004, A&A, 421, 491

Piconcelli E., et al., 2003, A&A, 412, 689

Pradas J., Kerp J., 2005, A&A, 443, 721.

Ptak A., et al., 2006, ApJ, 637, 147

Pounds K., et al., 1989, MNRAS, 240, 769.

Risaliti G., Maiolino R., Salvati M., 1999, ApJ, 522, 157

Risaliti G., et al., et al., 2005, ApJ, 623, L93

Risaliti G., et al., 2006, ApJ, 637, L17

Rosati P., et al., 2002, ApJ, 556, 667

Rowan-Robinson M., et al., 2005, AJ, 129, 1183

Seymour N., McHardy I.M., Gunn K., 2004, MNRAS, 352, 131

Severgnini P., et al., 2003, A&A, 406, 483

Severgnini P., et al., 2006, 2006, A&A, 451, 859

Silk J., Rees M., 1998, A&A, 331, L1

Simpson C., Rawlings S., Lacy M., 1998, MNRAS, 306, 828

Simpson C., 2005, MNRAS, 360, 565

Spergel D.N., et al., 2003, ApJS, 148, 175

Springel V., Di Matteo T., Hernquist L. 2005, MNRAS, 361, 776

Steffen A.T., et al., 2003, ApJ, 596, L23

Stern D., et al. 2005, ApJ, 631, 163

Streblyanska A., et al., 2004, Nucl.Phys.B(Proc. Suppl.) 132, 232

Sturm E., et al., 2005, ApJ, 629, L21

Struder L., et al., 2001, A&A, 356, L18

Szokoly G.P., et al., 2004, ApJS, 155, 271

Teerikorpi P., 2004, A&A 424, 73

Tozzi P., et al., 2006, A&A, 451, 457

Treister E., et al., 2004, ApJ, 616, 123

Treister E., Urry C.M., 2005, ApJ, 630, 115

Trump J.R., et al., 2006, ApJS, 165, 1

Turner T.J., et al., 1997, ApJS, 113, 23

Turner M.J.L., et al., 2001, A&A, 356, L27

Ueda Y., et al., 2003, ApJ, 598, 886

Wandel A., Peterson B.M., Malkan M.A., 1999, ApJ, 526, 579

Watson M., et al., 2001, A&A, 365, L51

Weisskopf M.C., et al., 2000, Proc. SPIE, 4012, 2

Williams R.J., Pogge R.W., Mathur S., 2002, AJ, 124, 3042

Williams O.R., et al., 1992, ApJ, 389, 157

Wolf C., et al., 2003, A&A 408, 499

Wolf C., et al., 2004, A&A, 421, 913

Yan L., et al., 2005, ApJ, 628, 604

Yu Q., Tremaine, S., 2002, MNRAS, 335, 965

Zehavi I., et al., 2002, ApJ, 571, 172

Zehavi I., et al., 2005, ApJ, 621, 22

Zheng W., et al., 2004, ApJS, 155, 73

Books

Krolik, J.H., "Active Galactic Nuclei", 1999, Princeton University Press

Rybicki G.B., Lightman A.P., "Radiative processes in astrophysics", 1979, Wiley-Interscience.

XMM-Newton calibration documents

Aschenbach B., et al., 2001, XMM-SOC-CAL-TN-0005

Arnaud K.A., 1996, Astronomical Data Analysis Software and Systems V, eds. Jacoby G. and Barnes J., p17, ASP Conf. Series volume 101.

Ehle M., et al., *XMM-Newton* Users Handbook, Issue 2.3, 08/07/2005.

Ghizzardi S., 2002, XMM-SOC-CAL-TN-0029

Gondoin P., 2000, *XMM-Newton* CCF release note, XMM-CCF-REL-2

Kirsch M., et al., 2005, XMM-SOC-CAL-TN-0018

Osborne J., 2001, SSC-LUX-TN-0059 v2.0
(http://xmmssc-www.star.le.ac.uk/pubdocs/ssc-lux-tn-0059.shtml)

Tody D., 1986, "The IRAF Data Reduction and Analysis System", Proc. SPIE, 627, 733